UNIVERSIDAD AUTONÓMA DE MADRID
FACULTAD DE CIENCIAS
Departamento de Biología Molecular

Characterization of P1 leader proteases of the *Potyviridae* family and identification of the host factors involved in their proteolytic activity during viral infection

Hongying Shan

Ph.D. Dissertation

Madrid 2018

UNIVERSIDAD AUTONOMA DE MADRID
Facultad de Ciencias
Departamento de Biología Molecular

Characterization of P1 leader proteases of the *Potyviridae* family and identification of the host factors involved in their proteolytic activity during viral infection

Hongying Shan

This thesis is performed in *Departamento de Genética Molecular de Plantas of Centro Nacional de Biotecnología* (CNB-CSIC) under the supervision of Dr. Juan Antonio García and Dr. Bernardo Rodamilans Ramos

Madrid 2018

Acknowledgements

First of all, I want to express my appreciation to thesis supervisors Bernardo Rodamilans and Juan Antonio García, who gave the dedicated guidance to this thesis.

I also want to say thanks to Carmen Simón-Mateo, Fabio Pasin, Raquel Piqueras, Beatriz García, Mingmin, Zhengnan, Wenli, Linlin, Ruiqiang, Runhong and Yuwei, who helped me and provided interesting suggestions for the thesis as well as technical support.

Thanks to the people in the greenhouse (Tomás Heras, Alejandro Barrasa and Esperanza Parrilla), *in vitro* plant culture facility (María Luisa Peinado and Beatriz Casal), advanced light microscopy (Sylvia Gutiérrez and Ana Oña), photography service (Inés Poveda) and proteomics facility (Sergio Ciordia and María Carmen Mena).

Thanks a lot to all the assistance from lab313 colleagues.

Thanks a lot to the whole CNB.

Thanks a lot to the Chinese Scholarship Council.

Thanks a lot to all my friends.

Thanks a lot to my family.

Madrid

20/03/2018

Index

CONTENTS

ABBREVIATIONS

Abbreviations

2, 4-D	2, 4-dinitrophenol
$2A^{pro}$	2A protease
$3C^{pro}$	3C protease
3'UTR	3' untranslated region
5´UTR	5' untranslated region
Φ	Empty vector
aa	Amino acid
ACN	Acetonitrile
AGO	Argonaute protein
AlkB	Alkylated DNA repair protein
bp	Base pair
BY-2	*Nicotiana tabacum* Bright Yellow -2
BYL	*Tobacco* BY-2 cell-free lysate
BYL-EP	BYL prepared from evacuolated protoplasts
BYL-EP(DN)	Denatured BYL prepared from evacuolated protoplasts
BYL-P	BYL prepared from protoplasts
cDNA	Complementary DNA
cEF	Concentrated eluted fractions
cFT	Concentrated Flow-through fraction
CI	Cylindrical inclusion
CID	Collision-induced dissociation
CHCA	α-Cyano-4-hydroxycinnamic acid
CP	Capsid protein
dpa	Days post agroinfiltration
dsRNA	Double-strand RNA
dsDNA	Double-strand DNA
DTT	Dithiothreitol
DUB	Deubiquitylating enzyme
emPAI	Exponentially modified protein abundance index
EF	Fractions eluted
ER	Endoplasmic reticulum
FI	Fluorescence intensity

FPLC	Fast protein liquid chromatography
FT	Flow-through
GFC	Gel filtration chromatography
GFP	Green fluorescence protein
GUS	ß-Glucuronidase
HC	Helper component
HEL	Helicase
HSP89.1	Heat shock protein 89.1
HSP70	Heat shock protein 70
IAA	Indole-3-acetic acid
ICTV	International Committee on Taxonomy of Viruses
IgG	Immunoglobulin G
KAc	Potassium acetate
kDa	Kilodalton
LC ESI-MS/MS	Nano Liquid Chromatography coupled to Electrospray Tandem Mass Spectrometry
Lpro	Leader protease
MALDI-TOF	Matrix Assisted Laser Desorption Ionization - Time of Flight
MMTS	Methyl methanethiosulfonate
MS medium	Murashige & Skoog basal salts with minimal organics
MT/MET	Methyltransferase
MW	Molecular weight marker of protein
NIapro	Nuclear inclusion A protein
NIb	Nuclear inclusion B protein
NS2	Nonstructural protein 2
NS3	Nonstructural protein 3
nt	Nucleotide
OD	Optical Density
ORF	Open reading frame
P2a	Polyprotein 2a
P2ab	Polyprotein 2ab
PCP	Papain-like cysteine protease domain
PCR	Polymerase chain reaction

PDS	Phytoene desaturase
pdTp	Deoxythymidine 3′, 5′-bisphosphate
PMF	Peptide mass fingerprint
pI	Isoelectric point
PR, Pro	Protease
pre-CP	Precursor of capsid protein
RbcL	RubisCO large subunit
RdRp	RNA-dependent RNA-polymerase
(+)RNA	Positive strand RNA
(-)RNA	Negative strand RNA
RRL	Rabbit reticulocyte lysate
RSS	RNA silencing suppression
RT-qPCR	Quantitative reverse transcription PCR
SDS-PAGE	Sodium dodecyl sulfate polyacrylamide gel electrophoresis
TCEP	Tris(2-carboxyethyl) phosphine
TEAB	Triethylammonium bicarbonate
TFA	Trifluoroacetic acid
Tudor-SN protein 1	Tudor- staphylococcal nuclease protein 1
UBA	Ubiquitin –associated protein
USP12	Ubiquitin specific protein 12
UV	Ultraviolet
VIGS	Virus-induced gene silencing
VPg	Genome-linked viral protein
WGE	Wheat germ extract
WG/GW	Tryptophan-glycine and/or glycine-tryptophan motifs

VIRUS CITED

Virus cited

AgMV	*Agropyron mosaic virus*
BlVY	*Blackberry virus Y*
BNYVV	*Beet necrotic yellow vein virus*
BStMV	*Brome streak mosaic virus*
BVDV	*Bovine viral diarrhea virus*
CaMV	*Cauliflower mosaic virus*
CBSV	*Cassava brown streack virus*
CVYV	*Cucumber vein yellowing virus*
EAV	*Equine arteritis virus*
EuRV	*Euphorbia ringspot virus*
FMDV	*Foot-and-mouth disease virus*
HEV	*Hepatitis E virus*
HIV	*Human immunodeficiency virus*
MCDV	*Maize chlorotic dwarf virus*
MLV	*Murine leukemia virus*
PLRV	*Potato leafroll virus*
PPV	*Plum pox virus*
PRSV	*Papaya ringspot virus*
PVY	*Potato virus Y*
RSV	*Rous sarcoma virus*
RTBV	*Rice tungro bacilliform virus*
RTSV	*Rice tungro spherical virus*
RYMV	*Rice yellow mottle virus*
SARS-CoV	*Severe acute respiratory syndrome coronavirus*
SeMV	*Sesbania mosaic virus*
SMoV	*Strawberry mottle virus*
SMV	*Soybean mosaic virus*
SPFMV	*Sweet potato feathery mottle virus*
SPMMV	*Sweet potato mild mottle virus*
TBSV	*Tomato bushy stunt virus*
TMV	*Tobacco mosaic virus*
ToMMV	*Tomato mild mottle virus*

TomRSV	*Tomato ringspot virus*
TriMV	*Triticum mosaic virus*
TRV	*Tobacco rattle virus*
TuMV	*Turnip mosaic virus*
TVMV	*Tobacco vein mottling virus*
TYMV	*Turnip yellow mosaic virus*
UCBSV	*Ugandan cassava brown streak virus*
WSMV	*Wheat streak mosaic virus*

SUMMARY

Summary

Viral endopeptidases play a key role in most viruses that use the polyprotein strategy for the production of their proteins. They are involved in several functions of the viral life cycle, such as replication, virion maturation or host range determination, thanks to a detailed modulation of their proteolytic activity. Their prominence during viral infection and their specificity in terms of proteolysis make them ideal antiviral targets as well as promising biotechnological tools.

The *Potyviridae* family is one of the largest families of plant-infecting viruses including ca. 200 species, which causes serious economic impact across the globe. As most plant viruses, this family, composed of seven monopartite and one bipartite genera, presents a positive single stranded RNA genome. Most viruses in the family encode P1 at the N-terminus, a serine protease with autocatalytic activity.

This thesis characterizes P1 proteins in the *Potyviridae* family as Type A or Type B based, among other features, in their requirement or not of plant factor(s) to perform their proteolytic activity. It highlights the relevance of Type A P1 proteins as host range determinants by studying *Plum pox virus* (PPV) P1 and *Cucumber vein yellowing virus* (CVYV) P1a processing capacities in the context of permissive and non-permissive hosts. It also demonstrates that a mutant virus carrying an N-terminal deletion on P1 is able to enhance PPV replication in cucumber plants, reinforcing the role of this protein´s region as negative regulator of the infection and the significance of P1 proteins in host range delimitation. In addition, the analysis of the proteolytic domains of other Type A P1 proteins, such as CVYV P1a and *Turnip mosaic virus* P1, identifies the C-terminal region as relevant for host factor(s) interaction.

As part of the characterization of leader proteases of the *Potyviridae* family, this thesis attempts to identify the host factor(s) involved in the proteolytic process of Type A P1 proteins following a biochemical approach based on the fractionation of an extract of *Nicotiana tabacum* cells. This method allowed identifying nine proteins as possible factors relevant for PPV P1 self-cleavage ability setting up the basis for the final identification of the host factor(s).

RESUMEN

Resumen

Las endopeptidasas virales desempeñan un papel clave en los virus que obtienen algunos de sus productos génicos a partir de poliproteínas. Están implicadas en diferentes funciones del ciclo viral, como son la replicación, la maduración del virión o la supresión de defensas de la planta, y contribuyen significativamente a la determinación del espectro de huésped, gracias a una modulación precisa de su actividad proteolítica. Su relevancia durante la infección viral, junto con su especificidad de corte las convierten tanto en atractivos objetivos de estrategias antivirales como en potenciales herramientas biotecnológicas.

La familia *Potyviridae* es una de los grupos de virus de plantas más extensos. Cuenta con alrededor de 200 especies, muchas de ellas con un enorme impacto socioeconómico a nivel mundial. Como la mayor parte de virus de plantas, los miembros de esta familia, que está compuesta por siete géneros con genoma monopartito y uno con genoma bipartito, presentan un genoma de RNA de polaridad positiva de hebra sencilla que se expresa por medio de poliproteínas. En la mayoría de los géneros de la familia, en el extremo amino-terminal de la poliproteína traducida desde el extremo 5' del RNA genómico se localizan una o varias serin proteasas denominadas P1, con actividad autocatalítica.

Esta tesis caracteriza las proteínas P1 de la familia *Potyviridae* como tipo A o tipo B basándose, entre otras características, en la existencia o no de requerimientos de un factor(es) de la planta para desarrollar su actividad proteolítica. Destaca la comprobación, mediante el estudio de las capacidades autocatalíticas de las proteínas P1 del *Plum pox virus* (PPV) y P1a del *Cucumber vein yellowing virus* (CVYV) en contextos de huéspedes permisivos y no-permisivos, de la relevancia de las proteínas P1 tipo A como determinantes del espectro de huésped viral. También demuestra que un virus mutante con una deleción del dominio amino terminal de P1 es capaz de potenciar la replicación del PPV en plantas de pepino, resaltando así el papel de esta región de la proteína como regulador negativo de la infección y reafirmando la importancia de las proteínas P1 en la discriminación de hospedadores. Además, el análisis de los dominios proteolíticos de otras proteínas P1 tipo A, como la P1a del CVYV y la P1 del *Turnip mosaic virus*, ha permitido identificar la región carboxilo-terminal como relevante en la interacción con el co-factor(es) del huésped.

Como parte de la caracterización de las proteasas líder de la familia *Potyviridae*, esta tesis intenta identificar el factor(es) de la planta implicados en el procesamiento proteolítico de las proteínas P1 tipo A siguiendo un abordaje bioquímico basado en el fraccionamiento de un extracto de células cultivadas *in vitro* de *Nicotiana tabacum*. Este método ha permitido identificar nueve proteínas como factores relevantes para el auto-procesamiento de la proteína P1 del PPV, sentando las bases para la identificación final de su co-factor(es) del huésped.

INTRODUCTION

I. Introduction

Viruses are ubiquitous and affect every organism in the biosphere (Moelling, 2012). Since the discovery of *Tobacco mosaic virus* (TMV) in the late 19th century (Zaitlin, 1998), more than 4400 viral species have been identified and categorized, a number that should increase dramatically when new data obtained from environmental metagenomics were incorporated into the ICTV classification system. According to this taxonomy, as of 2017, viruses can be classified in 8 orders comprising a total of 119 families. Most of the families, however, remain unassigned to an order, which reflects the diversity of these organisms and the difficulty to reconstruct their evolutionary histories (Simmonds *et al.*, 2017;Adams *et al.*, 2017). Viruses can also be classified based on the nature of their genome in RNA, DNA and retro-transcribing viruses; and RNA viruses can be subdivided into superfamilies based on RNA-dependent RNA-polymerase (RdRp) phylogenetic relationships: Picornavirus-like, Alphavirus-like and Flavivirus-like (Dolja & Koonin, 2011;Koonin, 1991;Goldbach *et al.*, 1991).

Viruses do not encode ribosomes, a feature that makes them fully dependent on the translation machinery of the host cell (Walsh *et al.*, 2013). With the exception of giant viruses (Wilhelm *et al.*, 2017), viruses share a reduced genome size and try to optimize a confined genetic space utilizing several strategies of alternative protein production (Firth & Brierley, 2012;Miras *et al.*, 2017): Internal ribosomal entry, leaky scanning, non-AUG initiation, reinitiation, frameshifting, readthrough or stop-carry on, are among these strategies. Another way that viruses follow to expand their protein endowment is to produce polyproteins that are further processed by proteases into smaller working units. This strategy ensures production of multiple components required for viral infection in a single molecule and at the same time saves space in the genome by using a single set of transcriptional and translational control elements. It also provides the option to yield partially processed protein products with specific activities, and to alter functionality of a particular protein in a controlled manner (Konvalinka *et al.*, 2015;Spall *et al.*, 1997). However, gene expression trough polyproteins heavily relies on proteases for its proper functioning and as so, these enzymes play a central role regulating the viral cycle.

Out of the 119 viral families, 26 include viruses infecting plants, with 11 orphan genera. Of these, there are 11 families of plant viruses and 1 unassigned genus that use the polyprotein strategy at some point to express all or part of their gene products (Table I.1, Figure I.1) (Adams et al., 2017). The largest family of plant viruses is *Geminiviridae*, whose members carry a single strand DNA genome, although the plant virome is heavily dominated by viruses with positive-strand RNA genomes, being *Potyviridae* the largest representative family of this class (Ivanov *et*

al., 2014;Revers & García, 2015). Among plant viruses there are also pararetroviruses and viruses with negative strand and double strand RNA genomes.

Viral endopeptidases share certain features that make them distinct from host proteases: i) they are smaller, ii) they present specific folds unique in the protease world, with amino acid sequences unrelated to other proteases of the same class, iii) they can adapt to multiple roles, and iv) they are very specific in their cutting requirements (Babe & Craik, 1997;Tong, 2002;Verdaguer *et al.*, 2014). This stringent specificity of viral proteases makes them ideal as biotechnological tools (Fernandez-Rodriguez & Voigt, 2016;Kim *et al.*, 2012;Tran *et al.*, 2017) and as targets for antiviral therapies (Shamsi *et al.*, 2016). Different drugs targeting proteases have been used successfully for treating animal viral infections (Anderson *et al.*, 2009;Clark *et al.*, 2013;Gable *et al.*, 2014;Raut *et al.*, 2015;Wu *et al.*, 2016), and have also reached moderate success in the plant world (García *et al.*, 1993;Gholizadeh *et al.*, 2005;Gutiérrez-Campos *et al.*, 2001;Gutierrez-Campos *et al.*, 1999;Habib, 2007;Kim *et al.*, 2016;Wen *et al.*, 2004).

In general, plant viral proteases have been understudied when compared to their animal counterparts in terms of processing regulation and structure, probably due to the relevant role that the latter play in human health. Plant viral proteases carry on multiple roles during viral infection independent of their protease activity: RNA silencing suppression (RSS), aphid transmission, systemic transport, viral accumulation, viral particle enhancement, etc. (Csorba *et al.*, 2015;Valli *et al.*, 2018;Liu *et al.*, 2009). As proteases, however, the primary role they play in viral infection is processing of and from viral polyproteins. But there is more in this protease activity that just acting as peptide cutters. Polyprotein processing is not an all-or-nothing process in which all products are separated at the same time with perfect efficiency. Cleavage of the polyprotein into functional units is essential for viral survival and it is a highly modulated process. Its regulation modifies the timing and place of the final products as well as the possible accumulation of intermediate products, which can play distinct roles in the life cycle. In addition, processing of host proteins can also implement functionalities of these viral proteases.

Following is an overview of different roles that lie behind the proteolytic activity of plant virus proteases emphasizing their relevance during viral infection. As the object of study of this thesis, P1 leader proteases of the family *Potyviridae* will be presented in a separate section.

Table 1. Plant virus proteases

Family	Group[1]	Proteases	Type	MEROPS[2] Family	Action	Suggested specific function
Potyviridae	(+)ssRNA Picorna-like	P1a-like	Serine	S30	*cis*	Replication Counterdefense Host range definition
		P1b-like	Serine	S30	*cis*	Counterdefense
		HC	Cysteine	C6	*cis*	Aphid transmission Counterdefense Virion assembly
		NIapro	Cysteine	C4	*cis/trans*	Replication Host range definition Superinfection exclusion
		P2-1	Cysteine	C6	*cis*	Unknown
Secoviridae	(+)ssRNA Picorna-like	Pro	Cysteine	C3	*cis/trans*	Replication Counterdefense
Luteoviridae	(+)ssRNA Picorna-like	Protease	Serine	S39	*cis*/*trans*	Replication
Sobemovirus[3]	(+)ssRNA Picorna-like	P1	Unknown	Unclassified	*cis*	Counterdefense
		Pro	Serine	S39	*cis*/*trans*	Replication
Tymoviridae	(+)ssRNA alpha-like	PRO	Cysteine	C21	*cis*/*trans*	Replication Counterdefense
Closteroviridae	(+)ssRNA alpha-like	L/P/L1/L2	Cysteine	C42	*cis*	Systemic movement Host range definition Superinfection exclusion
Betaflexiviridae	(+)ssRNA alpha-like	PRO	Cysteine	C23	*cis*/*trans*	Replication
Benyiviridae	(+)ssRNA alpha-like	PCP	Cysteine	C36	*cis*/*trans*	Replication Counterdefense
Endornaviridae	dsRNA alpha-like	CRR?	Cysteine?	Unclassified	*cis*/*trans*?	Unknown
Pseudoviridae	(+)ssRNA Retrovirus	PR	Aspartic	A11	*cis*/*trans*	Virion maturation?
Metaviridae	(+)ssRNA Retrovirus	PR	Aspartic	A2	*cis*/*trans*	Virion maturation?
Caulimoviridae	dsDNA Pararetrovirus	PR	Aspartic	A3	*cis*/*trans*	Virion maturation

[1] Classification of RNA viruses based on (Dolja & Koonin, 2011)
[2] MEROPS classification of proteases (Rawlings *et al.*, 2014)
[3] This genus is unassigned to a family.

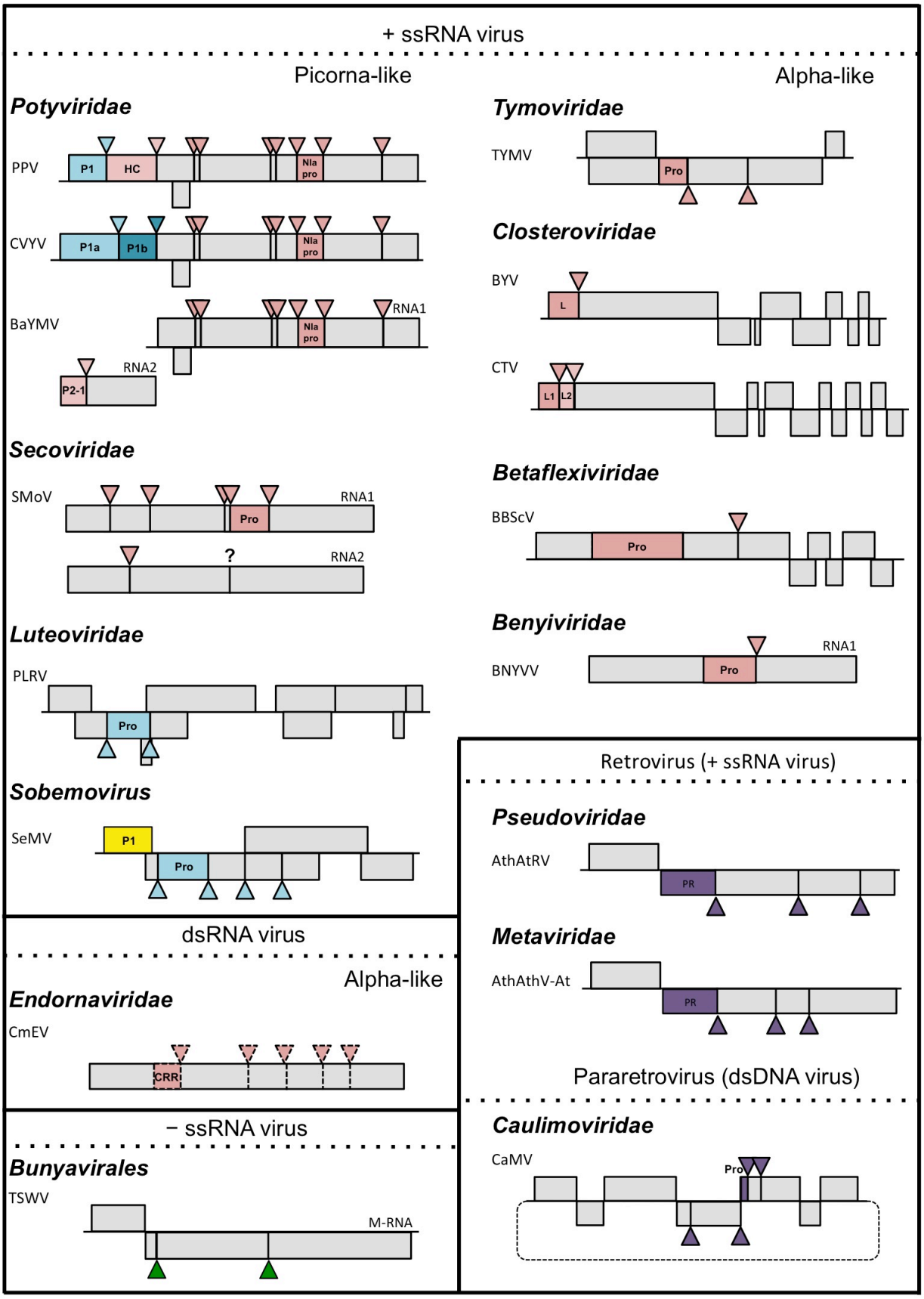

Figure I.1. Schematic representation of plant viruses and their proteolytic cleavage sites. Triangles represent cleavage sites of endopeptidases. Colors of the triangles match the colors of the corresponding endopeptidases: pink

for cysteine, blue for serine, purple for aspartic, yellow for unknown and dark green for plant proteases; only genomes, or sub-genomes encoding polyproteins subject to proteolytic cleavage are depicted. For each family, a representative species covering the different endopeptidases is depicted. Scale of the genome map is maintained only within each viral species. Dotted lines used in the *Endornaviridae* family indicate that processing is only theoretical. Question mark indicates that the way of processing is unknown. PPV: *Plum pox virus*, *Potyvirus*; CVYV: *Cucumber vein yellowing virus*, *Ipomovirus*; BaYMV: *Barley yellow mosaic virus*, *Bymovirus*; SMoV: *Strawberry mottle virus*, Unassigned; PLRV: *Potato leafroll virus*, *Polerovirus*; SeMV: *Sesbania mosaic virus*, *Sobemovirus*; TYMV: *Turnip yellow mosaic virus*, *Tymovirus*; BYV: *Beet yellow virus*, *Closterovirus*; CTV: *Citrus tristeza virus, Closterovirus;* BBScV: *Blueberry scorch virus*, *Carlavirus*; BNYVV: *Beet necrotic yellow vein virus*, *Benyvirus*; CeMV: *Cucumis melo alphaendornavirus*, *Alphaendornavirus*; TSWV: *Tomato spotted wilt orthotospovirus*, *Orthotospovirus*; AthAtRV: *Arabidopsis thaliana AtRE1 virus*, *Pseudovirus*; AthAthV-At: *Arabidopsis thaliana athila virus*, *Metavirus*; CaMV: *Cauliflower mosaic virus*, *Caulimovirus*.

I.1. Functions of plant viral proteases

I.1.1 Replication

Key to viral infection is genome replication. It takes place at specific sites in the cell, compartments termed viral factories, in which multiple viral and plant factors required for replication are concentrated (Heinlein, 2015). Involvement of viral proteases in these factories has been demonstrated for some cases, but information is not always available. For animal viruses, the role that endopeptidases play in regulating replication is well established (Racaniello, 2001;Yost & Marcotrigiano, 2013;Sawicki & Sawicki, 1994;Vasiljeva *et al.*, 2003;Rausalu *et al.*, 2016). Information is scarcer in the case of plant infecting viruses.

The *Potyviridae* is a family of positive-strand RNA viruses that belongs to the picornavirus-like supergroup. It comprises 8 genera and presents the highest protease variety among plant viruses, coding in their genomes up to five different proteases with varied specificities [P1 (Type A and B), Helper component (HC), NIa protein (NIapro), P2-1 (HC-like)] (Revers & García, 2015;Adams *et al.*, 2005a;Adams *et al.*, 2005b;Rodamilans *et al.*, 2013). A hallmark of the picorna-like viruses, besides a conserved RdRp, is the presence of a 3C-like protease in charge of polyprotein processing. For the *Potyviridae*, this is NIapro. Indeed, this is the best characterized plant viral protease, functionally and structurally, that modulates replication by polyprotein processing (Carrington & Dougherty, 1987). NIapro is a chymotrypsin-like cysteine protease that acts *in cis* and *in trans* and it is involved in the generation of intermediate (such as P3-6K1, CI-6K2, 6K2-NIa) and final products at different stages of infection. These products are implicated in the formation of the replication complex and its anchoring to, and release from, ER-derived membranes (Restrepo-Hartwig & Carrington, 1994;Beauchemin *et al.*, 2007;Merits *et al.*, 2002;Cui & Wang, 2016;García *et al.*, 2014;Riechmann *et al.*, 1995;Schaad *et al.*, 1997).

Belonging to the picornavirus-like supergroup and sharing equivalent proteases are the *Secoviridae* (Thompson *et al.*, 2014), and *Luteoviridae* (Li *et al.*, 2000;Prüfer *et al.*, 1999) families and the *Sobemovirus* genus (Sõmera *et al.*, 2015;Satheshkumar *et al.*, 2004). In the *Secoviridae* family, studies with waikavirus *Rice tungro spherical virus* (RTSV) (Thole & Hull, 1998), nepovirus *Tomato ringspot virus* (TomRSV) (Wang *et al.*, 1999;Wang & Sanfaçon, 2000) and *Strawberry mottle virus* (SMoV) (Mann *et al.*, 2017), have characterized the viral protease (Pro) and their cleavage sites, but still much is left to know about the specific involvement of these proteases in viral replication. Same is true for the *Luteoviridae* family. The serine protease encoded by the ORF1 of the polerovirus *Potato leafroll virus* (PLRV) is able to act *in cis* and *in trans* and separates the membrane anchoring portion, the protease and the genome-linked viral protein (VPg) domains; whether this is part of a regulatory mechanism for viral replication is still unknown (Li *et al.*, 2007). Sobemoviruses express two versions of a polyprotein, from ORF2a and ORF2b, having different C-terminus. N-terminal common part includes membrane anchor domain, the protease Pro, and VPg. Polyprotein 2a (P2a) C-terminal part codes for P10 and P8 proteins. Polyprotein 2ab (P2ab) codes for RdRp and is originated by ribosomal frameshift. Studies with *Sesbania mosaic virus* (SeMV) indicate that serine protease performs differently in P2a and in P2ab (Nair & Savithri, 2010a;Nair & Savithri, 2010b). In the first case, processing occurs at the predicted sites separating all components from the polyprotein. However, in the latter case, processing of VPg from RdRp is not fulfilled even though the protease and cleavage sequence are conserved in P2a and P2ab. This points to a regulatory process in protease activity that might have an influence in replication considering the inhibitory effect observed *in vitro* that VPg has over the polymerase when present at its N-terminus. In addition, mutational analysis of cleavage sites indicated that all sites at p2a/p2ab are essential for viral replication, and the products are only functional when released at the site of replication (Govind *et al.*, 2012) reinforcing the modulatory role of the protease.

Another example of plant viral protease involved in replication comes from the *Tymoviridae* family that belongs to the alphavirus-like supergroup. *Turnip yellow mosaic virus* (TYMV) encodes a papain-like cysteine protease, termed PRO (Rozanov *et al.*, 1995;Lombardi *et al.*, 2013). Involvement of PRO in replication comes from two different lines of evidence: i) the processing ability of the protease to act *in cis* and *in trans* similarly to the proteases of rubiviruses and alphaviruses, which share a similar polyprotein structure (Jakubiec *et al.*, 2007;Jakubiec *et al.*, 2004) and ii) its deubiquitination activity (Camborde *et al.*, 2010;Chenon *et al.*, 2012). TYMV is the type member of the genus *Tymovirus*, a single positive-strand RNA spherical virus that produces two overlapping ORFs from a single RNA. One of them encodes a

polyprotein of 206 kDa that contains sequence domains of methyltransferase (MT), PRO, helicase (HEL) and RdRp. PRO was shown not only to separate RdRp from the rest of the polyprotein, but also process HEL in a secondary event. This and the ability of PRO to act *in trans* tighten the evolutionary relationship of this virus to rubiviruses and alphaviruses, and, as it occurs in these animal viruses, it is likely that temporal regulation of polyprotein processing controls the synthesis of different RNA species (negative- and positive-strands). Whether the specific cleavage observed in TYMV also shuts off the synthesis of negative-strand RNA is still unknown (Jakubiec et al., 2007). In addition to this, TYMV PRO is a functional ovarian tumour-like deubiquitylating enzyme (DUB) and this activity helps PRO to modulate viral replication by stabilizing the viral polymerase preventing degradation by the ubiquitin-proteosome system (Bailey-Elkin *et al.*, 2014;Camborde et al., 2010;Chenon et al., 2012;Jupin *et al.*, 2017).

The alphavirus-like supergroup does not maintain a conserved protease in all members as the picornavirus-like does. In this way, the *Closteroviridae* family, although sharing in ORF1a the MET, HEL organization followed by RdRp in ORF1b, does not encode a protease that acts *in trans* to process these products, but present a leader proteinase(s) with autocatalytic activity (Dolja, 1994). On the other hand, some members of the *Betaflexiviridae* family, do encode in ORF1 similar MET, PRO, HEL, RdRp domains as *Tymoviridae* viruses, although there is little information regarding polyprotein processing and no data regarding involvement of PRO in replication (Lawrence *et al.*, 1995;Foster & Mills, 1992). Similar lack of information is encountered in the *Benyiviridae* family. Its most studied member, *Beet necrotic yellow vein virus* (BNYVV), encodes a papain-like cysteine protease domain (PCP) (Hehn *et al.*, 1997), and it has been hypothesized that it might act as a DUB to favor RdRp transcription (Pakdel *et al.*, 2015), similarly to the mode of action of the PCP domain of *Hepatitis E virus* (HEV), although in the latter case, PCP acted as a DUB to counteract cellular antiviral pathways (Karpe & Lole, 2011).

I.1.2 Viral counterdefense

Sometimes, when a protease potentiates a positive effect on replication it is not due to a specific role in this viral process, but it is the consequence of an indirect effect caused by an enhanced ability of the virus to scape plant defenses. Thus, proteases could be considered as having a counterdefense role instead of a role in viral replication. For instance, if RdRp degradation is considered as part of plant defense, TYMV PRO activity as DUB, can be viewed not in terms of modulating replication, but as a counterdefense mechanism (Camborde et al., 2010;Chenon et al., 2012;Lombardi et al., 2013;Jupin et al., 2017).

Considering DUBs in terms of protection against host defenses is something well established in the animal viral world. Examples can be found among viruses of the order *Nidovirales* such as the coronavirus *Severe acute respiratory syndrome coronavirus* (SARS-CoV) or the arterivirus *Equine arteritis virus* (EAV) that use this strategy of interfering with the innate immune signaling pathway through the DUB activity of their cysteine proteases (Clementz *et al.*, 2010;van Kasteren *et al.*, 2013). Same is true for viruses of the order *Picornavirales* such as the aphtovirus *Foot-and-mouth disease virus* (FMDV) and its L^{pro} leader protease (Wang *et al.*, 2011a;Wang *et al.*, 2011b). In all these cases, however, although the counterdefense activity is well documented, it appears that the DUB and the protease activity are not strictly interrelated (Jupin et al., 2017).

Probably, the best characterized proteases acting as viral counterdefense barriers by degrading host proteins are the ones from the *Picornaviridae* family. Thus, FMDV L^{pro} not only disrupts the interferon signaling pathway through its deubiquitinase activity but also cleaves eIF4G shutting off host cap-dependent translation and downregulating Type I interferons (Chase & Semler, 2012;Guarné *et al.*, 1998;Liu *et al.*, 2015). Moreover, FMDV produces, as the rest of the members of the *Picornaviridae* family, $3C^{pro}$, a protease that is in charge of processing the different elements of the polyprotein acting *in cis* and *in trans*, and also degrades several host proteins in order to potentiate viral transcription and translation (Sun *et al.*, 2016). In the same family, rhinoviruses and enteroviruses produce another protease termed $2A^{pro}$, which also develops these degrading functions (Chase & Semler, 2012;Seipelt *et al.*, 1999).

Taking these activities into account, it is reasonable to ask the question as to whether the 3C-like proteases of plant picorna-like viruses perform similar host degrading activities to counteract plant defenses or not. In the case of NIapro, the 3C-like protease of the *Potyviridae*, there has not been described specific host proteins affected by its catalytic activity and, only recently, a study was published describing possible interacting partners in plants (Martínez *et al.*, 2016). However, it cannot be discarded that NIapro might be processing more proteins than the viral ones taking into account its demonstrated ability to act on proteins with an engineered target sequence (Rohila *et al.*, 2004b;Cesaratto *et al.*, 2016) or even on proteins with a naturally occurring target cleavage site, such as the amyloid-ß peptide (Han *et al.*, 2010;Kim et al., 2012). In this line, NIapro from *Potato virus Y* (PVY) acts as elicitor of the hypersensitive response mediated by the gen *Ry* in potato, and its protease activity, likely acting on a host factor, appears to be involved in this eliciting response (Mestre *et al.*, 2000;Mestre *et al.*, 2003). More recent studies have described a role of potyviral NIapro in enhancing aphid transmission and suggested

that this role might be related to its ability to degrade vacuolar defense proteins (Casteel *et al.*, 2014;Bak *et al.*, 2017).

Some newly published reports add more information to the scarce available data about activities of 3C-like proteases related with defense and couterdefense responses. The RNA silencing suppressor R78 of the waikavirus *Maize chlorotic dwarf virus* (MCDV) is cleaved by Pro, raising the possibility that this cleavage might have some influence in R78 silencing suppression activity over the course of the infection (Stewart *et al.*, 2017). Moreover, NIapro of the tritimovirus *Wheat streak mosaic virus* (WSMV) contributes to prevent superinfection by related viruses, and it has been suggested that the protease activity of this protein is required for the superinfection exclusion (Tatineni & French, 2016).

I.1.3 Virion maturation

A good example of a viral protease directly involved in virion formation is togavirin from viruses of the genus *Alphavirus*. Structurally related to chymotrypsin-like serine proteases, togavirin is the actual core protein. It self-processes from the polyprotein precursor, binds viral RNA, and assembles into the capsid (Krupovic & Koonin, 2017). Apart from this versatile endopeptidase, the role of proteases in virion maturation has been well studied for animal retroviruses such as *Human immunodeficiency viru*s (HIV), *Rous sarcoma virus* (RSV) or *Murine leukemia virus* (MLV), amongst others (Konvalinka et al., 2015). In these viruses, cleavage of viral polyproteins at specific sites and in an orderly fashion is crucial for transforming the immature shell into an active infectious particle. *Pseudoviridae* and *Metaviridae* are two viral families that include plant retroviruses (Eickbush & Jamburuthugoda, 2008;Peterson-Burch & Voytas, 2002;Wright & Voytas, 2002), but there is not much information regarding the regulation of the proteolytic processing. More data is available about the *Caulimoviridae*, the single family of plant pararetroviruses (Torruella *et al.*, 1989). The genome of all replication-competent retroviruses consists of structural, replication and envelope proteins (gag, pol, env) (Marmey *et al.*, 2005). The protease (PR), an aspartate peptidase with no homology to other viral proteases, is generally included in the pol domain. *Caulimoviridae*, the only family of plant viruses with dsDNA genomes, encode the gag-pol core, but unlike retroviruses, lacks an integrase, which is not required because the caulimoviral DNA is not integrated in the host chromosome. The type virus of the family is *Cauliflower mosaic virus* (CaMV), a member of the *Caulimovirus* genus. The capsid protein (CP) of this virus is produced as a precursor (pre-CP) with N- and C-terminal extensions. CP is involved in virion assembly, packaging of viral RNA and delivery of the genome to the nucleus. Processing of the CP

extensions is thought to regulate these functions. The N-terminal extension of CP appears to be involved in keeping the pre-CP in the cytoplasm and may operate as an anchoring domain for the initiation of viral assembly, similarly to what occurs to HIV viral matrix protein (Champagne *et al.*, 2004). Virion maturation is completed by removal of the first 76 aa and about 40 aa from the C terminus by the viral aspartic proteinase (Champagne et al., 2004;Karsies *et al.*, 2002). The fact that pre-CP is excluded from the nucleus would assure that only mature virions, containing the genomic DNA, enter in the nucleus (Karsies et al., 2002). Studies done with another plant pararetrovirus, the badnavirus *Rice tungro bacilliform virus* (RTBV), showed that its aspartic protease cuts independently of plant-specific host factors since it retained its proteolytic activity in baculovirus (Laco *et al.*, 1995) and bacteria (Marmey et al., 2005). In the case of animal retroviruses, PR is expressed in an inactive monomeric form and needs to dimerize to get into active conformation in which each unit contributes with an aspartate to the active site. Proper Red-ox environment is likely to also play a role in PR activation (Konvalinka et al., 2015;Ingr *et al.*, 2003). Based on active site comparison, it is anticipated that PR of *Caulimoviridae* also acts as dimers (Torruella et al., 1989). Its activation requirements are still pending of further investigation.

I.1.4 Host range definition

Plant viruses have definite host ranges, which in some cases are very narrow. The complex network of interactions between plant and virus that needs to be established in order for the infection to progress makes it difficult for the virus to have broad host spectrum. In terms of viral proteases, the best examples of host range modulation come from the P1 proteases in the *Potyviridae* family, and will be discussed in the corresponding section. But inside this family, NIapro also has been described to play a role in host range determination. In *Papaya ringspot virus* (PRSV) a single amino acid substitution in this chymotrypsin-like protease allows to shift hosts from cucurbits to papaya, although the specific involvement of the protease activity of NIapro in this effect is only a possibility (Chen *et al.*, 2008). More direct evidence of the involvement of the protease activity of NIapro in host range determination comes from work performed with *Plum pox virus* (PPV) (Calvo *et al.*, 2014). This study showed that alternative adaptation to *Nicotiana* and *Prunus* hosts was determined, not by peculiarities of the NIapro sequence, but by differences in the NIapro target sequence placed between 6K1 and CI, suggesting modulation of NIapro processing at this site in a host-specific manner.

I.1.5 Proteolytic activity-unrelated functions

The small size of the genome of plant RNA viruses forces the proteins from these viruses to gather multiple functions. This is best exemplified by the potyviral protein HC (Valli et al., 2018). HC is a cysteine proteinase whose first identified function was to aid in aphid transmission of viral particles (de Mejia *et al.*, 1985). However, the main function of the potyviral HC appears to be suppressing antiviral RNA silencing (Anandalakshmi *et al.*, 1998;Kasschau & Carrington, 1998), and an independent function of HC in the correct assembly of potyviral virions has been more recently reported (Valli *et al.*, 2014). Interestingly, all these HC functions do not rely on its proteolytic activity, illustrating how proteolysis-related and -unrelated roles can concur in a single viral protein.

Viral proteinases with functions that appear not to be related with their proteolytic activity are not restricted to the family *Potyviridae*. The self-cleaving leader proteinases of viruses of the *Closteroviridae* family are a good example of this. These proteinases are involved in virus accumulation, systemic transport, host range expansion and virus superinfection exclusion, but all these roles appeared to be independent of their protease activities (Atallah *et al.*, 2016;Liu et al., 2009;Peng *et al.*, 2003;Peng *et al.*, 2001). Contrary to what was observed for the leader proteinase of FMDV, the closterovirus proteases show no DUB activity and have not been described to be involved in further processing of host or viral proteins.

I.1.6 Functions still uncharacterized

It is well established the notion that viral proteases are not just proteolytic machines acting without proper modulation of time and/or space. Much effort has been put into defining what are these extra roles and characterizing the different mechanisms of action and their peculiarities. Involved in regulating replication, virion maturation, host range determination or even displaying a more active role as viral counterdefense barriers, proteases, when present, are essential in practically all aspects of the viral cycle. However, still there are many proteinases from plant viruses for which information about the integration of its enzymatic activity in the infection process is still unavailable. Viruses of the family *Endornaviridae* are a fine example. These viruses have been understudied probably because they do not usually cause any noticeable damage on their hosts. They present a monocistronic RNA genome that encodes a large polyprotein, but there are only hints about how this polyprotein is processed (Roossinck *et al.*, 2011;Sabanadzovic *et al.*, 2016). The case of P1 of the sobemovirus *Rice yellow mottle virus* (RYMV) is another good example of a viral protease with a puzzling role (Weinheimer *et al.*, 2010). RYMV P1, a protein with RNA silencing suppression activity, is expressed as a mature

protein, rather than as part of a protein precursor; however, in experimental conditions it displays self-cleaving activity able to precisely remove engineered C-terminal extensions. Maintaining a function that seems to be superfluous raises the possibility that this protease, and by similarity other leader proteases, might have an extra biological function that remains elusive. In addition, control of gene expression by proteolytic processing of protein precursors not only relies on viral proteinases. For instance, host aspartyl proteases are in charge of the processing of the primary product of the M genomic RNA of plant viruses of the order *Bunyavirales* to yield two mature glycoproteins (Li *et al.*, 2015;Shi *et al.*, 2016;Whitfield *et al.*, 2005).

I.2. Leader proteases in the *Potyviridae* family

Potyviridae is a plant viral family whose members produce an enormous ecological and economic impact affecting herbaceous and woody species across the globe. These viruses bear a single positive-strand RNA and form filamentous, flexuous particles. The *Potyviridae* family comprises 195 species; 2 of them are unassigned to any genera and the rest are divided into seven monopartite genera and one bipartite genus. Most of the species (160) belong to the *Potyvirus* genus (Wylie *et al.*, 2017) (Adams et al., 2017). All viruses in the family use the polyprotein strategy for protein production and in all species the central and carboxy-terminal regions of the polyprotein follow a conserved organization, P3-6K1-CI-6K2-VPg-NIapro-NIb-CP, and are processed by NIapro. Exceptionally, the ipomoviruses *Cassava brown streack virus* (CBSV) and *Ugandan cassava brown streak virus* (UCBSV) and the potyvirus *Euphorbia ringspot virus* (EuRV) encode a HAM protein between NIb and CP. In addition to this, all species present in the middle region of P3 a GA_6 conserved motif that causes polymerase slippage giving rise to two extra products, P3N-PIPO (+A) and P3N-ALT (-A) (Olspert *et al.*, 2015;Rodamilans *et al.*, 2015;Hagiwara-Komoda *et al.*, 2016). Bymoviruses, the bipartite potyvirids, have this same genomic organization (from P3 to CP including the GA_6 motif) in RNA1 (Adams et al., 2005a;Adams et al., 2005b;Revers & García, 2015).

I.2.1 Proteolytic diversity of potyvirids

As mentioned early in the chapter, *Potyviridae* is the family of plant viruses with the highest diversity in terms of proteases. All viruses exhibit the hallmark 3C-like protease of the Picorna-like viruses, NIapro. But the proteolytic diversity is a reflection of the highly variable N-terminal part of the polyprotein of the different viruses of the family (Figure I.2). Potyviruses and rymoviruses present at this region a P1 serine protease followed by cysteine protease, HC. Both

present autocatalytic activity and HC is in charge of RSS, among other mentioned functions (Valli et al., 2018). Besides, potyviruses infecting sweet potato, such as *Sweet potato feathery mottle virus* (SPFMV) present an enlarged P1 that bears the slippage sequence GA_6. This originates an extra ORF and the production of P1N-PISPO, a protein with RSS activity (Mingot *et al.*, 2016;Untiveros *et al.*, 2016). Tritimoviruses and poaceviruses also follow the P1-HC tandem array, but in these genera, P1 and not HC is the protease in charge of RSS (Tatineni *et al.*, 2012;Young *et al.*, 2012). Similar N-terminal configuration of the polyprotein can be found in the genus *Brambivyrus* (Susaimuthu *et al.*, 2008). However, in this case, its sole member, *Blackberry virus Y* (BlVY), encodes an oversized P1 carrying an AlkB domain in the middle region followed by a short HC. The RSS ability of these proteins is undetermined. Macluraviruses lack a P1 leader protease and encode a reduced version of HC, of unknown functionalities (Kondo & Fujita, 2012). Ipomoviruses are the most variable in terms of N-terminal polyprotein arrangements. Some viruses, such as *Tomato mild mottle virus* (ToMMV) present a P1-HC configuration similar to tritimo- and poaceviruses with P1 carrying RSS activity, followed by HC (Abraham *et al.*, 2012). *Sweet potato mild mottle virus* (SPMMV) also follows this scheme, but its P1 is enlarged and the RSS activity is based on the WG/GW AGO-binding motif (Giner *et al.*, 2010). Other viruses in the genus, such as *Cucumber vein yellowing virus* (CVYV), present two P1 proteases in tandem (P1a and P1b) with the second one in charge of RSS activity (Valli *et al.*, 2008;Valli *et al.*, 2007;Valli *et al.*, 2006). *Cassava brown streak virus* (CBSV) and *Ugandan cassava brown streak virus* (UCBSV) do not encode HC either, but only present a single P1 copy with RSS activity (Mbanzibwa *et al.*, 2009). Finally, the only bipartite genus in the family *Potyviridae*, the *Bymovirus*, carries two RNAs. RNA1, as mentioned, encodes a polyprotein that starts at P3 without leader proteases; RNA2 encodes two proteins, P2-1 and P2-2, being P2-1 a cysteine HC-like protease with autocatalytic activity (Kashiwazaki *et al.*, 1991;Kashiwazaki *et al.*, 1990).

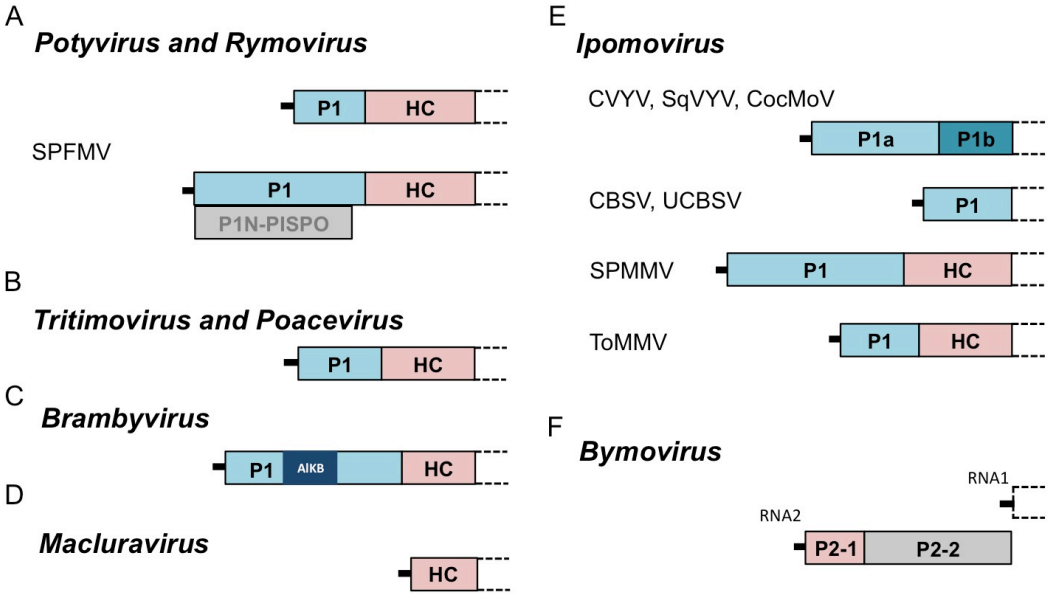

Figure I.2 Variable N-terminal part of the polyprotein of the different viruses of *Potyviridae* family. Colors of endopeptidases: pink for cysteine, blue for serine. Scale of the N-terminal genome map is maintained only within each viral species. Dotted lines indicate omitted viral genomes. Black square represents AlKB domain in P1 protein. SPFMV: *Sweet potato feathery mottle virus*; CVYV, *Cucumber vein yellowing virus*; SqVYV: *Squash vein yellowing virus*; CocMoV: *Coccinia mottle virus*; CBSV*: Cassava brown streak virus*; UCBSV: *Ugandan cassava brown streak virus*; SPMMV: *Sweet potato mild mottle virus*; ToMMV: *Tomato mild mottle virus*.

I.2.2 Functions of P1 proteases

The best example of the diversity of P1 leader proteases is CVYV. Initially this virus was considered to have a large P1 with a single proteolytic activity (Janssen *et al.*, 2005). Detailed *in silico* analyses allowed finding of an extra serine protease motif in the middle of this large P1 and further experiments helped to characterize the presence of two proteins at the N-terminal region of CVYV, P1a and P1b (Valli et al., 2008;Valli et al., 2007;Valli et al., 2006). P1b presented RSS activity and, similar to HC, this function was independent of its proteolytic activity. Phylogenetically, this P1 leader protease was related to P1s of tritimo- and poaceviruses and similar to other P1 proteins of ipomoviruses. On the other hand, P1a presented similarities to P1s of potyviruses and rymoviruses in terms of phylogeny and pI, but, most relevant, displayed the need for a plant factor(s) to perform its proteolytic activity (Valli et al., 2006;Rodamilans et al., 2013;Verchot *et al.*, 1992), an activity that is essential for virus infectivity (Verchot & Carrington, 1995b). This requirement for a plant factor(s) is probably the reason why Type A proteases display a different arrange of functions not found in Type B P1s. Early works characterizing the proteolytic activity and the host factor requirements of these proteases did not clarify what functions, other than autocatalytic activity was performing P1, which was dubbed a

"mysterious" protein in a review (Rohozkova & Navratil, 2011). But several studies conducted in the last ten years certainly help elucidate some of the roles that Type A P1 proteases might have in the viral life cycle of the *Potyviridae*.

An important feature of the potyviral P1s is being a determinant of host range. The comparison of two PPV isolates, which differed in their reciprocal capacity of infecting woody and herbaceous hosts, showed the relevance of P1 among other viral proteins for host adaptation (Salvador *et al.*, 2008a). Similarly, analyses of PPV chimeras including P1 sequences of *Tobacco vein mottling virus* (TVMV) and of virus variants with different biological properties sorted from a single PPV isolate also pointed towards the implication of P1 in host range definition (Salvador *et al.*, 2008b;Maliogka *et al.*, 2012). All these works show how relevant is P1 in terms of host spectrum definition, but not necessarily implicate the protease activity of P1 in this role. More direct evidence of the involvement of P1-mediated proteolytic processing in compatibility with the host comes from works performed with P1a of CVYV and P1 of PPV, both Type A P1 proteases. In these study, it was shown that one of the factors limiting the infection of a chimeric PPV carrying CVYV P1a instead of PPV P1 in *Nicotiana benthamiana* was likely the incompatibility of CVYV P1a with a host co-factor required for its protease activity (Carbonell *et al.*, 2012).

Recently, the leader protease P1 of the *Potyvirus* genus has also been assigned a role in controlling viral replication. Work performed with PPV P1 showed that the N-terminal part of this *cis*-acting serine proteinase, the most variable region, acts as a negative regulator of P1 self-processing, modulating in this way potyviral replication (Pasin *et al.*, 2014b). Removal of the N-terminal part of P1, not only makes the protein co-factor independent, but also potentiates viral replication at early times of infection emphasizing the regulatory role of this protein in the potyviral life cycle. The way PPV P1 is modulating replication through host factor interactions resembles the way of action of NS2 protease of animal virus *Bovine viral diarrhea virus* (BVDV) (Lackner *et al.*, 2004;Lackner *et al.*, 2006). In this pestivirus, NS2 protease modulates replication indirectly by downregulating NS2-NS3 processing. Similarly, PPV P1 modulates P1HC processing and indirectly affects viral replication.

I.3. Objectives

Regulation of viral encoded endopeptidases is complicated, affected by different layers of modulation in time and space. This thesis intends to improve knowledge about *Potyviridae*, one of the largest families of plant viruses, considering how relevant proteolytic regulation is for the life cycle of these viruses. The thesis focuses on P1 leader proteases, present in almost all members of the family, as their proteolytic activity is essential for viral infectivity, representing a good example of viral endopeptidase subject to modulation.

The specific objectives of this thesis are:

1. To classify P1 proteins in *Potyviridae* family.

2. To elucidate the relevance of P1 proteolytic activity in host range definition.

3. To identify and characterize the host factor(s) determining PPV P1 self-cleavage ability.

MATERIALS AND METHODS

II. Materials and methods

II.1 Virus and bacterial strains

Brome streak mosaic virus (BStMV) and UCBSV were purchased from DSMZ plant virus collection (PV-1052 and PV-0912, respectively).

Escherichia coli strain DH10B was used for cloning of plasmids and vector amplifications and strain DB3.1, containing the *gyr*A426 allele, which renders the strain resistant to the toxic effects of the ccdB gene, was used for amplification of plasmids carrying the toxin gene. *Agrobacterium tumefaciens* C58C1 harboring a disarmed pTiB6S3 plasmid was used for leaf infiltration for viral inoculation and transient expression assays.

II.2 Plant hosts, trees and callus

Transient expression of proteins and viral infectivity assays by agroinfiltration were carried out in *Nicotiana benthamiana* and *Cucumis sativus* Albatroz RZ F1 (Rijk Zwaan Iberica, Almería, Spain). Another set of viral infections was performed in seedlings of *Prunus avium* "Pontavium", *Prunus domestica* "Brompton" and *Prunus persica* "GF305" (Pépinières Lafond, Valreas, Cedex, France). The germination process and growing conditions of these trees were as described (Calvo et al., 2014). All above mentioned plants were grown in greenhouse maintained at a 16 h light/8 h dark photoperiod and a temperature range of 18-23 ºC.

Nicotiana tabacum cv. BY-2 (Bright Yellow - 2) callus was gifted by Pilar Fontanet (Center for Research in Agriculture Genomics, Spain). BY-2 suspension cells were cultured in the dark under 26-27 ºC, 160 rpm in MS medium [4.43 g/L of Murashige & Skoog basal salts with minimal organics (Sigma-Aldrich), 3% sucrose, 1 µg/ml thiamine-HCl, 200 µg/ml KH_2PO_4, 1 mg/ml $MgSO_4$-$7H_2O$, 0.2 µg/ml 2,4-D, adjusted to pH 5.8 with KOH] (Murota *et al.*, 2011). Solid BY-2 cell callus was maintained in MS medium with agar 0.8%. The cell suspension was regenerated every week by transferring 2.5 ml saturated BY-2 culture into 47.5 ml MS medium. Solid callus was renewed every 4 weeks transferring 4 or 5 pieces of 1-2 mm well-growing callus to a new MS plate.

II.3 Agroinfiltration

N. benthamiana, *C. sativus*, *P. avium*, *P. domestica* and *P. persica* were infiltrated with *A. tumefaciens* carrying the indicated plasmids. Appropriate *Agrobacterium* cultures were sedimented by centrifugation at 5000 g under room temperature for 10 minutes and resuspended in induction buffer (10 mM MES, 10 mM $MgCl_2$, 0.15 mM acetosyringone, pH 5.6). Cells were

incubated in this medium for ~3 h at room temperature. Mixes were prepared at the indicated OD_{600} in each test (detailed in results) and applied with a syringe to the underside of 2-3 leaves of 4-week-old plants with 5-6 real leaves. In each experiment, two independent *A. tumefaciens* cultures containing the same set of constructs were delivered in parallel. Fluorescence intensity and protein accumulation were determined at the indicated time points.

II.4 Fluorescence imaging and quantification

Green Fluorescence Protein (GFP) fluorescence images were acquired by confocal laser scanning microscopy (TCS SP5 or SP8 system, excitation laser at 488 nm and emission bandwidth of 505–549 nm; Leica), by an epifluorescence microscope (MZ FLIII, GFP3 filter; Leica) or by placing leaves on a blue light transilluminator (Safe Imager, Invitrogen) and taking photographs with a Nikon D3X digital camera. GFP fluorescent intensity quantification was carried out placing individual 5.0 mm-diameter leaf discs in a black 96-well plate (Nunc) filled with 50 µl water/well (to limit sample dehydration) and measured by a monochromator-based plate reader (Infinite M200, Tecan Group) (Pasin *et al.*, 2014a).

II.5. DNA plasmids

II.5.1 Viral constructs

A full-length cDNA copy of a PPV isolate adapted to *Nicotiana* (Riechmann et al., 1995), tagged with sGFP(S65T) (Chiu *et al.*, 1996) and inserted in the pSN-ccdB binary plasmid was reported (Pasin et al., 2014b), as well as the viral cDNA vector into which the PPV P1-HC sequence was replaced by PPV P1S(Ser259Ala), PPV P1Pro(Δ163 amino acid) (Pasin et al., 2014b) and CVYV P1a-P1b (Carbonell et al., 2012). TVMV cDNA clone was described and provided by E. Rodriguez-Cerezo (Centro Nacional de Biotecnología, Spain) (Moreno *et al.*, 1999) *Turnip mosaic virus* (TuMV) cDNA clone was previously described (Sánchez *et al.*, 1998) and kindly provided by Fernando Ponz (Centro de Biotecnología y Genómica de Plantas, Spain). cDNA clones from *Tobacco rattle virus* (TRV) RNA 1 and RNA 2 (pTRV1 and pTRV2, respectively) were bought from TAIR – ABRC (Liu *et al.*, 2002a;Liu *et al.*, 2002b). Cloning details of the newly prepared viral constructs are described below.

pSN-PPV P1a; PPV P1 sequence was replaced by CVYV P1a and the sequence corresponding to five extra residues of CVYV P1b and a GlySerGly linker. This sequence was inserted upstream of the PPV HC sequence. Full-length P1a was amplified from pDONR207-P1Stop (Pasin et al., 2014b) with primers 2095_F and 2096_R. pONE (Pasin et al., 2014b) was used as

2063_R and part of the HC coding sequence with primer pair 2097_F and 1930_R. These three PCR products were used as templates for overlapping PCR using primers 1986_F and 1910_R. Inserts and *Sfa*AI-digested pSN-ccdB were mixed at a molar proportion 2:1 and used in Gibson assembly reaction (Gibson *et al.*, 2009).

pSN-PPV P1aS; PPV P1 sequence was replaced by the same sequence as in *pSN-PPV P1a*, but in this case CVYV P1a carried a mutation giving rise to a Ser484Ala mutation. P1a full-length fragment was amplified into 2 fragments by PCR with primers 2095_F and 2237_R, and 2238_F and 2096R, which incorporated the nucleotide changes required to create the Ser484Ala mutation. Overlapping PCR was performed to get full-length P1a-S484A. The rest of the cloning process was the same as in pSN-PPV P1a.

pSN-PPV P1aNIa; PPV P1 sequence was replaced by CVYV P1a, five extra residues of CVYV P1b, a GlySerGly linker and the PPV NIa protease cleavage site SNVVVHQ↓ADE. This sequence was inserted upstream of the PPV HC sequence. PCR amplification was similar as in pSN-PPVP1a, except that amplification of HC allowed introduction of the sequence coding for the NIa cleavage site, using primer pair 2098_F and 1930_R. The rest of the cloning process was the same as for pSN-PPV P1a.

pSN-PPV P1ProΔGDD; pSN-PPV P1Pro and a fragment amplified from pLONG-NIbΔGDD (Gallo, 2017) by PCR with primers 67_F and 2264_R were digested with XbaI/BamHI and combined with T4 DNA ligase (Thermo Fisher) for overnight ligation.

pSN-PPV-5´BD; pICPPV-5'BD-GFP (Salvador et al., 2008a) was digested with ScaI/BamHI/SacI (SacI is used to digest unused fragment to avoid self-assembly), and the fragment of about 8 Kb was mixed with ScaI/BamHI-digested pSN-ccdB (treated with shrimp alkaline phosphatase) and T4 DNA ligase (Thermo Fisher) for overnight ligation.

pTRV2-VIGS plasmids; Total RNA from *N. benthamiana* was extracted using TRIzol (Thermo Fisher). cDNA was prepared with Superscript III, following manufacturer´s instructions.

Group 1 [ß-Glucuronidase (GUS), phytoene desaturase (PDS) and TUDOR-SN 1]. PDS (1 fragment) and TUDOR-SN 1 (2 fragments) sequences were amplified from *N. benthamiana* cDNA using appropriate primers (Table II.1). GUS (1 fragment) was amplified from pIC-PPVnkGUS (Lucini, 2004) using appropriate primers (Table II.1). PCR fragments and pTRV2 were digested with EcoRI/XhoI and combined with T4 DNA ligase (Thermo Fisher) for overnight ligation.

Group 2 (HSP89.1, HSP70, USP12, UBA and unknown protein). HSP89.1 (1 fragment), HSP70 (2 fragments), USP12 (4 fragments), UBA (1 fragment) and unknown protein (1 fragment)

sequences were amplified from *N. benthamiana* cDNA using appropriate primers (Table II.1). pTRV2 was digested with XbaI/XmaI (BamHI and SacI digestion were also included to reduce background colonies). Digested plasmid and PCR-amplified fragments were combined using a Gibson assembly kit (NEBuilder® HiFi DNA Assembly Master Mix).

All above ligation mixtures were transformed in *E. coli* DH10B competent cells. Colonies from pSN-PPV P1a, pSN-PPV P1aS, and pSN-PPV P1aNIa were checked using colony PCR with primers 1985_F and 1597_R. Colonies from pSN-PPV P1Pro Δ GDD were checked using colony PCR with primers 2950_F and 2264_R. Colonies from pTRV2-VIGS constructs were checked using colony PCR with primers X261_F and X262_R. Positive clones were grown in liquid medium and plasmids purified. Plasmid DNA of pSN-PPV P1a, pSN-PPV P1aS and pSN-PPV P1aNIa were further checked by HindIII/SpeI double digestion and sequencing (Macrogen). Plasmid DNA of pSN-PPV P1Pro Δ GDD was further checked by BamHI/XbaI double digestion and sequencing (Macrogen). Plasmid DNA of pTRV2-VIGS constructs were further checked by HindIII/DraIII double digestion and sequencing (Macrogen). Primers, listed in Table II.1, were synthesized by Sigma-Aldrich.

Table II.1. List of primers used in the study.

Primers	Sequence (5'→3')
Q26_R	TTTCGTCTTCTACTTGCTAACGGACTCTTGCTCCAATTCCTACAGATAAC
Q27_F	ATCCTACACGATGGCAGTGAAAAC
Q28_R	TCTTCTACTTGCTAACGGACTCTTG
Q29_F	GCAACACCTCTACAATGTGATAACCGACTGATGCAGGAACTGGAGCAAGC
Q30_F	CTCTACAATGT GATAACCGACTGATG
Q31_R	CGCCGTCTTTTGGGATGAA
67_F	GGATGAAGTTTGCTGG
697_F	GGGGACAAGTTTGTACAAAAAAGCAGGCTGGAAAATATAAAAACTCAACACAAC
698_R	GGGGACCACTTTGTACAAGAAAGCTGGGTCTCCAACCAGGTATGTTTTCATATTTG
699_F	GGGGACAAGTTTGTACAAAAAAGCAGGCTACTCTGACCCAGGCAAACAATTTTGG
700_R	GGGGACCACTTTGTACAAGAAAGCTGGGTTATGGTTCTTCCTTAAGCACCTGC
1630_F	CAGACGAAGGCAGCAGCATTG
1632_R	GTTTTTTTTTTTTTTTTGTCTCTTGC
1985_F	AGAGGATTGACGTGATAACATGG
2009_R	TAATTCCTCGAGTCACTTTGGGGCTTGGGCATC
2063_R	CATCTTGACTTGCAGTAAATTTGGTAG
2172_F	CTATATAAGGAATAATACGACTCACTATAGGGAAAATATAAAAACTCAAC

2173_R	GAACATTTCTCAATGCTGCTGCCTTCGTCTGGCTAAGCGAGAACATGTGAAAATTG
2174_F	CCTTAATTTCTCTACCAAATTTACTGC
2229_F	TCTCTACCAAATTTACTGCAAGTCAAGATGGTTAGAATGTCCGAGGCATC
2231_F	ACCAAATTTACTGCAAGTCAAGATGTCAAAGACAGTCACAAAGAGAGAAG
2232_F	CTACCAAATTTACTGCAAGTCAAGATGGCAAAAAGTCTCATAAACACATA
2233_F	TCTACCAAATTTACTGCAAGTCAAGATGGCAGCTATTCACCAGGATAATG
2236_F	TCTACCAAATTTACTGCAAGTCAAGATG_AAGACGAGAACCGAGTATTCTG
2237_R	CTACCGTTATCAAACCGGCAGTTCCAGGA
2238_F	CTGGAACTGCCGGTTTGATAACGGTAGG
2264_R	GTGAGCGGATAACAATTTCAC
2312_F	CCTTAATTTCTCTACCAAATTTACTGCAAGTCAAGATGCAGCAGACGCACGTTAAGC
2313_R	GAACATTTCTCAATGCTGCTGCCTTCGTCTGGCTAAATCATTAACTGTCCACTTGC
2316_F	GGGGACAAGTTTGTACAAAAAAGCAGGCTTCATGAGTAGCAAGAGGCGCATGGC
2317_R	GGGGACCACTTTGTACAAGAAAGCTGGGTTAACCAATCTCGTATTCCTTAAGTTC
2318_F	ATGCAGCAGACGCAC
2320_R	CTATAATACTCAATCCGCTC
2358_F	GCAGGGGTTGTGTTGC
2359_R	GCAACACAACCCCTGC
2461_F	GGGGACAAGTTTGTACAAAAAAGCAGGCTGGAAATTAAATGATCATACGC
2462_R	GGGGACCACTTTGTACAAGAAAGCTGGGTTACCCCACGCGGTACTCTTGAC
2463_R	GGGACCACTTTGTACAAGAAAGCTGGGTTAGTATTGTAGGTAGTCGTCAGC
2464_F	GGGGACAAGTTTGTACAAAAAAGCAGGCTGGACCATGAGTGAGGAGTTGCAACCATTG
2465_R	CAATGGTTGCAACTCCTCACTCATTCTGATCTTGTTGTGATTC
2467_F	GGGGACAAGTTTGTACAAAAAAGCAGGCTGGACCATGCCCACACGGTACAGGGG
2500_F	CCTTAATTTCTCTACCAAATTTACTGCAAGTCAAGATGCCCACACGGTACAGGGG
2501_R	GAACATTTCTCAATGCTGCTGCCTTCGTCTGGCTATGGAGCTGGGTCCGCACATTTG
2502_F	CTTTTGGCACAGCTGGAG
2503_R	CTCCAGCTGTGCCAAAAG
2749_F	TCTCTACCAAATTTACTGCAAGTCAAGATGTCTGTTCGGAGTTTGGTGAAGG
2800-F	TCTCTACCAAATTTACTGCAAGTCAAGATGAGAACCGAGTATTCTGTTCG
2801_F	TCTCTACCAAATTTACTGCAAGTCAAGATGGAGTATTCTGTTCGGAGTTTG
2802_F	TCTCTACCAAATTTACTGCAAGTCAAGATGCGGAGTTTGGTGAAGGAAATTG
2803_F	TCTCTACCAAATTTACTGCAAGTCAAGATGTTGGTGAAGGAAATTGGAAAAAC
2815_F	TCTCTACCAAATTTACTGCAAGTCAAGATGGCAGCAGTTACATTCGC
2818_R	GAACATTTCTCAATGCTGCTGCCTTCGTCTGGCTATACCTATCTAGTATCTGCACTG
2821_F	TCTCTACCAAATTTACTGCAAGTCAAGATGTGCAAGATGAACGACCAAGG
2822_F	TCTCTACCAAATTTACTGCAAGTCAAGATGCAAGGAGTTGACATGTTGACAC
2823_F	TCTCTACCAAATTTACTGCAAGTCAAGATGCGATCCCTGGTTAAGATTTTC.
2870_F	TCTACCAAATTTACTGCAAGTCAAGATGCCATTCCCTCATATTATTG

2871_F	TCTACCAAATTTACTGCAAGTCAAGATGATGAAAAAGAAGGTGGTGTTTAC
2876_F	TCTCTACCAAATTTACTGCAAGTCAAGATGGCTACAATTCATGGATTGCATG
2877_F	TCTCTACCAAATTTACTGCAAGTCAAGATGGCTAGCGAGTTCCAATTAAGAG
2878_R	GGGGACCACTTTGTACAAGAAAGCTGGGTCCTAAAAGTCAATTTTATCTTTCTC
2879_R	GGGGACCACTTTGTACAAGAAAGCTGGGTCCTAGTCAATTTTATCTTTCTCATC
2880_R	GGGGACCACTTTGTACAAGAAAGCTGGGTCCTAAATTTTATCTTTCTCATCTGC
2881_F	TCTCTACCAAATTTACTGCAAGTCAAGATGGCAGCAACAATGATCTTTGG
2882_R	GAACATTTCTCAATGCTGCTGCCTTCGTCTGGCTA CAAGTCGTGTAAGAAGCGAAT
2897_F	TCTACCAAATTTACTGCAAGTCAAGATGCGGTGCTTCAAAGAAAGCTCTG
2898_F	TCTACCAAATTTACTGCAAGTCAAGATGATTGAAAAGAAGCTCAGG
2899_F	TCTACCAAATTTACTGCAAGTCAAGATGGGTCCGGATGCAATAGTC
2950_F	AAATTCAGAGGTGGTTGGGAC
2962_F	ACCAAATTTACTGCAAGTCAAGATGGCTGCGATGTCCGAGGCATCACTAC
3022_F	ATACAAGGAATAATACGACTCACTATTAGGAAAAATAAATATGACATAAG
3023_R	TGTGTTTTTTTCTTAATAATCCTATCTTACTACTCATACTTATTATATTC
3029_F	GTCCTGGATGGGCTGGAGCAATC
3030_R	GATTGCTCCAGCCCATCCAGGAC
3081_F	CGAGTACGTAGGATCCATTTAAATACTAGTCCAGTACGCACGATTC
3082_R	CATCTTGACTTGCAGTAAATTTGGTAG
3083_R	CTTTTGCTGTGGGCAAATCCACATACTTCGAATCACCTG
X79-C1_R	GCCTTCATCTGGATATGAGCTTCACCTAGTGGATTATCTCATTGCTCTGT
X79-C3_R	GCCTTCATCTGGATATGAGCTTCACCTATATCTCATTGCTCTGTTCTTTA
X79-C5_R	GCCTTCATCTGGATATGAGCTTCACCTAATTGCTCTGTTCTTTAGAAACC
X261_F	TTGTTAGATAATGGTTTGGTGGTC
X262_R	TGAtCCTAAAACTTCAGACACGG
V-GUS_F	CGGGAATTCCGATGCGGTCACTCATTAC
V-GUS_R	CGGCTCGAGGGTTTGTGGTTAATCAGGAAC
V-PDS_F	CGGGAATTCTGGGAGTTCCTGTGATAAATG
V-PDS_R	CGGCTCGAGGTGTACAACGCTAATTCAGCG
V-Tudor-1_F	CGGGAATTCCAAAAAGAAGAGGTGAAGGTAAC
V-Tudor-1_R	CGGCTCGAGAAGCTCACTTAGACGGTAAGC
V-Tudor-2_F	CGGGAATTCCCTCTGGACAACTATATATCCAGGAG
V-Tudor-2_R	CGGCTCGAGCACGCTCTAAGGCCTCCC
V-HSP89_F	GATTCTGTGAGTAAGGTTACCTAATTCTGCAGAAGCCAAGAAGGAAGGAG
V-HSP89_R	CCGTAGTTTAATGTCTTCGGGACATGCGAAGAATCTCTCGTGATACAT
V-HSP70-1_F	GATTCTGTGAGTAAGGTTACCTAATTCTCCCATCTTGGAGGTGAGGAC
V-HSP70-1_R	CCGTAGTTTAATGTCTTCGGGACATGCCAGCAGCACCATAAGCAACTG
V-HSP70-2_F	GATTCTGTGAGTAAGGTTACCTAATTCTGAAACTGCTGGAGGAGTGATG
V-HSP70-2_R	CCGTAGTTTAATGTCTTCGGGACATGCGGTGTTCCTCATGTTGTAGG

V-USP12-1_F	GATTCTGTGAGTAAGGTTACCTAATTCTGTAAAGATGATATACTCCTTTTC
V-USP12-1_R	CCGTAGTTTAATGTCTTCGGGACATGCGCTAACTCCAGACAGAAATC
V-USP12-2_F	GATTCTGTGAGTAAGGTTACCTAATTCTCAAGGAGGAAGTTGCAAAAG
V-USP12-2_R	CCGTAGTTTAATGTCTTCGGGACATGCCAAATTTTATTTCCTCAAAAAG
V-USP12-3_F	GATTCTGTGAGTAAGGTTACCTAATTCTGACAGATATGAGTTTCCTTTAG
V-USP12-3_R	CCGTAGTTTAATGTCTTCGGGACATGCGCTCTTTCTTCAGCCTTACCC
V-USP12-4_F	GATTCTGTGAGTAAGGTTACCTAATTCTCTCCTACAAACCCAATACCAT
V-USP12-4_R	CCGTAGTTTAATGTCTTCGGGACATGCCGTAGATTTAAAGTCCACATT
V-UBA_F	GATTCTGTGAGTAAGGTTACCTAATTCTGCTAGGGCTATCCGTGCACTTC
V-UBA_R	CCGTAGTTTAATGTCTTCGGGACATGCCGCAATGCTATTATACGTTTTC
V-Unknown_F	GATTCTGTGAGTAAGGTTACCGGATTCTTCGATGAGGATAAG
V-Unknown_F	CCGTAGTTTAATGTCTTCGGGACATGCTTGTCTCTCATCTTCCCTCAAG
V-Chorismate_F	GATTCTGTGAGTAAGGTTACCTAATTCTCAAGGTTGAAGTACCTAACAC
V-Chorismate_R	CCGTAGTTTAATGTCTTCGGGACATGCGCAGCAATCATCTTCTCTGC
V-tRNA_F	GATTCTGTGAGTAAGGTTACCTAATTCTCAGATCAACATCGTGGTTGG
V-tRNA_R	CCGTAGTTTAATGTCTTCGGGACATGCGCTCTCTCTGATGTTGTTCAC

II.5.2 Transient expression constructs

Transient expression vectors pSN.5 P1HC, pSN.5 P1SHC (expressing PPV P1HC and PPV P1S259AHC sequences, respectively) and the intermediate clone pSN2-ccdB have been already described (Pasin et al., 2014b) as was the pMDC32-NIaPro plasmid expressing the protease domain of the PPV NIa protein (Maliogka et al., 2012). *A. tumefaciens* C58C1 strains carrying the p35S:GFP (Haseloff, 1997) and pBIN61-p19 (Voinnet *et al.*, 2003) plasmids were kindly provided by Prof. D. Baulcombe (University of Cambridge, UK). Descriptions of newly generated transient expression constructs are detailed below:

pSN.5 P1aNIa; pSN.5-ccdB was digested with XbaI/BstBI and combined with SpeI/BstBI-digested pSN-PPV P1aNIa and T4 DNA ligase (Thermo Fisher Scientific) for overnight ligation.

pSN.5 P1a; pSN.5-ccdB was digested with XbaI/BstBI and combined with SpeI/BstBI-digested pSN-PPV P1a and T4 DNA ligase for overnight ligation.

pSN.5 P1aS; pSN.5-ccdB was digested with XbaI/BstBI and combined with SpeI/BstBI-digested pSN-PPV P1aS and T4 DNA ligase for overnight ligation.

pSN.5 P1Pro; pSN2-ccdB backbone was digested with XbaI/BstBI and combined with SpeI/BstBI-digested pSN-PPV P1Pro and T4 DNA ligase for overnight ligation.

pSN.5 P1aPro; Δ395P1aHC fragment was amplified from pSN-PPV P1a by PCR with primers 2236_F and 3083_R which contains a BstBI site. A second fragment including 35S promoter

and PPV 5'UTR was amplified with primers 3081_F and 3082_R. Both fragments were combined with XbaI/BstBI-digested pSN2-ccdB backbone and T4 DNA ligase for overnight ligation.

pSN.5 P1aSPro; Δ395P1aSHC fragment was amplified from pSN-PPV P1aS(S4848A) by PCR with primers 2236_F and 3083_R. The rest of the cloning process was the same as in pSN.5 P1aPro.

pMDC32-BStMV P1; Total RNA was extracted from BStMV-infected tissue using TRIzol (Thermo Fisher) and then, cDNA was prepared with Superscript III (Thermo Fisher), following manufacturer´s instructions. This cDNA was used as template for PCR amplification of full-length P1 with primers 2318_F and 2320_R. This PCR product carrying the gateway recombination sites was introduced first in pDONR207 and then in pMDC32 using BP and LR mixtures, respectively (Thermo Fisher).

pMDC32-BStMV P1S; BStMV cDNA was used as template for PCR amplifications of two P1 fragments with primers 2318_F and 2359_R and 2358_F and 2320_R. PCR products carrying nucleotides conferring a Ser355Ala mutation and gateway recombination sites, were joined by overlapping PCR and introduced first in pDONR207 and then in pMDC32 using BP and LR mixtures, respectively.

pMDC32-BStMV HC; BStMV cDNA was used as template for PCR amplification of the full-length HC sequence with primers 2316_F and 2317_R. This PCR product, carrying gateway recombination sites, was introduced first in pDONR207 and then in pMDC32 using BP and LR mixtures, respectively.

pMDC32-BlVY 5'UTRP1HC; BlVY 5´UTR-P1HC sequence was amplified from plasmid pUC57-BlVY, carrying the first 3000 nt of BlVY, (provided by Ioannis E. Tzanetakis, University of Arkansas System, United States) by PCR with primers 2461_F and 2462_R. This PCR product, carrying gateway recombination sites, was introduced first in pDONR207 and then in pMDC32 using BP and LR mixes, respectively.

pMDC32-BlVY P1; P1 sequence was amplified from pUC57-BlVY by PCR with primers 2467_F and 2463_R. This PCR product, carrying gateway recombination sites, was introduced first in pDONR207 and then in the pMDC32 using BP and LR mixtures, respectively.

pMDC32-BlVY HC; HC sequence was amplified from pUC57-BlVY by PCR with primers 2464_F and 2462_R. This PCR product, carrying gateway recombination tails, was introduced first in pDONR207 and then in pMDC32 using BP and LR mixtures, respectively (Thermo Fisher).

pP1HCTAP; P1HC sequence was amplified from pIC-PPVnkGFP (Fernández-Fernández *et al.*, 2001) by PCR with primers 697_F and 698_R. This PCR product, carrying gateway recombination sites, was introduced first in pDONR207 and then in *pCTAPi* (Rohila *et al.*, 2004a) using BP and LR mixtures, respectively.

pTAPHC; HC was amplified from pIC-PPVnkGFP (Fernández-Fernández et al., 2001) by PCR with primers 699_F and 700_R. This PCR product, carrying gateway recombination sites, was introduced first in pDONR207 and then in *pNTAPi* (Rohila et al., 2004a) using BP and LR mixes, respectively.

Standard molecular cloning methods were used performing PCR reactions with Phusion High-Fidelity DNA Polymerase (New England BioLabs) for the preparation of all plasmids and constructs. The accuracy of all newly built plasmids was verified by restriction digestion analysis and DNA sequencing (Macrogen) of PCR-generated inserts. Primers were synthesized by Sigma-Aldrich and sequences are listed in Table II.1**.**

II.6 DNA constructs to be expressed by *in vitro* transcription and translation

As described in (Pasin et al., 2014b), for the preparation of transcripts to be translated *in vitro* for PPV P1 proteolytic activity testing, three fragments were prepared and joined by overlapping PCR. PPV 5'UTR, including T7 promoter at its N-terminus, and 3'UTR fragments were obtained from pSN-PPV ΔP1-AAG and pIC-PPVnkGFP by PCR amplification with primers 1985_F/2063_R and 1630_F/1632_R, respectively. The middle fragment corresponding to PPV P1 and the first 97 amino acids of HC (HC-97) followed by a stop codon was amplified from pSN-PPV with primers 2174_F/2173_R. Overlapping PCR was performed with primers 2172_F/1632_R. This procedure was maintained for the production of the rest of the constructs in the study varying the PCR templates and the primers used. Details are shown in Table II.2. The integrity of PCR-generated DNA constructs was confirmed by gel electrophoresis, and constructs were used directly in *in vitro* transcription reactions.

Table II.2. Middle fragments prepared to build transcription constructs

Middle fragment	Template	Primers-F/R
T7_P1	pSN-PPV	2174_F/2173_R
T7_P1-S	pSN-PPVP1S	2174_F/2173_R
T7_P1Pro	pSN-PPV	2229_F/2173_R
T7_P1Pro-S	pSN-PPVP1S	2229_F/2173_R

T7_ΔP1pro-C1	pSN-PPV	2229_F/X79-C1_R
T7_ΔP1pro-C3	pSN-PPV	2229_F/X79-C3_R
T7_ΔP1pro-C5	pSN-PPV	2229_F/X79-C5_R
T7_P1bSTAP	pGG5S6N-P1bSTAP*	2172_F/2009_R
T7_ΔP1b-C1	pSN-PPVP1aP1b	2877_F/2878_R
T7_ΔP1b-C2	pSN-PPVP1aP1b	2877_F/2879_R
T7_ΔP1b-C3	pSN-PPVP1aP1b	2877_F/2880_R
T7_P1-A15	pSN-PPV	2233_F/2173_R
T7_P1-A48	pSN-PPV	2232_F/2173_R
T7_P1-S80	pSN-PPV	2231_F/2173_R
T7_P1-R90	pSN-PPV	2897_F/2173_R
T7_P1-I100	pSN-PPV	2898_F/2173_R
T7_P1-P120	pSN-PPV	2899_F/2173_R
T7_P1-P140	pSN-PPV	2870_F/2173_R
T7_P1-M155	pSN-PPV	2871_F/2173_R
T7_P1ProAA	pSN-PPV	2962_F/2173_R
T7_P1ProAA-S	pSN-PPVP1s	2962_F/2173_R
T7_P1a	pSN-PPV P1a	2172_F/2173_R
T7_P1a-K396	pSN-PPV P1a	2236_F/2173_R
T7_P1a-R398	pSN-PPV P1a	2800_F/2173_R
T7_P1a-E400	pSN-PPV P1a	2801_F/2173_R
T7_P1a-S402	pSN-PPV P1a	2749_F/2173_R
T7_P1a-R404	pSN-PPV P1a	2802_F/2173_R
T7_P1a-L406	pSN-PPV P1a	2803_F/2173_R
T7_TuMV P1	pUC19-TuMV	2815_F/2818_R
T7_TuMV P1-C218	pUC19-TuMV	2821_F/2818_R
T7_TuMV P1-Q223	pUC19-TuMV	2822_F/2818_R
T7_TuMV P1-R230	pUC19-TuMV	2823_F/2818_R
T7_TVMV P1	pMDC32-TVMV	2881_F/2882_R
T7_BlVY P1	pUC57-BlVY	2500_F/2501_R
T7_BlVY P1S**	pUC57-BlVY	2500_F/2501_R
T7_BstMV P1	pMDC32-BstMV	2312_F/2313_R
T7_BstMV P1S**	pMDC32-BstMV	2312_F/2313_R
T7_UCBSV P1	pLX-UCBSV	3022_F/ 3023_R

T7_UCBSV P1S**	pLX-UCBSV	3022_F/ 3023_R

* Plasmid was reported (Rodamilans et al., 2013)
** Serine mutations in P1 proteins of BlVY, BStMV and UCBSV were introduced with primers 2502-F/2503_R, 2358-F/2359_R and 3029-F/3030_R, respectively, followed by overlapping PCR with the primers specified in the table.

II.7 *In vitro* transcription and translation assays

In vitro cleavage assays were performed as reported (Pasin et al., 2014b). RNA transcripts were synthesized from the PCR-generated DNA templates described in the section II.6 using the HiScribe T7 High Yield RNA Synthesis Kit (NEB) and purified by organic extraction/ ammonium acetate precipitation. Quality and amount were assessed by spectrometry using a NanoDrop apparatus (Thermo Fisher Scientific) and gel electrophoresis, and final concentration adjusted to 200 ng/µL. *In vitro* translation was carried out in the presence of a mixture of L-[^{35}S] methionine and L-[^{35}S] cysteine (PerkinElmer) using the wheat germ extract (WGE) and rabbit reticulocyte lysate (RRL) systems (Promega). Samples were resolved in 10%-15% glycine-SDS-PAGE, depending on the size of the targeted proteins, and the radioactive signals were captured by phosphorimaging.

II.8 Western blot assays

Plant tissue was ground in a mortar under liquid nitrogen or powdered with a SPEX SamplePrep 2010 Geno/Grinder®. Crude extracts were prepared by homogenization in cracking buffer (125 mM Tris-HCl, 2% SDS, 6 M urea, 5% β-mercaptoethanol, 10% glycerol, 0.05% bromophenol blue, pH 7.5) using a mass:volume ratio of 1:2. Proteins were separated by 10%-15% glycine - SDS-PAGE depending on the size of the targeted products and electroblotted onto nitrocellulose membranes as reported (Pasin et al., 2014b). Anti-GFP monoclonal antibody (clones 7.1 and 13.1, Roche) and anti-PPV CP and anti-PPV HC rabbit sera were used as primary antibodies for protein detection; horseradish peroxidase conjugated sheep anti-mouse IgG (GE Healthcare) or goat anti-rabbit IgG (Jackson) were used as secondary antibodies. For signal quantification, chemiluminescence was acquired in a ChemiDoc XRS imager (BioRad) and analyzed with ImageJ software (Schneider *et al.*, 2012).

II.9 RT-qPCR analysis

Total RNA was extracted with FavorPrep Plant Total RNA Mini kit (Favorgen). Strand-specific quantification of PPV RNA was done with four biological replicates per condition using tagged cDNA primers in the RT step as described (Pasin et al., 2014b). Briefly, equal amounts of DNAseI-treated total RNA were used for cDNA synthesis using Superscript III (Invitrogen) and primer Q26_R or Q29_F (Table II.1) to transcribe cDNA from positive and negative PPV genomes, respectively. cDNA was treated with 0.5 U/µl Exonuclease I (Thermo Fisher) and purified with MinElute PCR Purification Kit (QIAGEN). Technical triplicate qPCR reactions were prepared using HOT FIREPol EvaGreen qPCR Mix Plus (Solis BioDyne) in 96-well optical plates and run in a 7500HT Fast Real-Time PCR System (Applied Biosystems). Primer pairs Q27_F/Q28_R and Q30_F/Q31_R (Table II.1) were used for positive and negative genome quantifications, respectively. The amount of target RNA in the analyzed samples was estimated by absolute quantification using external DNA standard curve. Quantification was done relative to the average value of PPV-P1Pro for the positive and the negative strands, respectively.

II.10 Sequence, phylogenetic and data analyses

Jalview 2.9 was used to visualize the potyviral P1 sequences, and sequence alignments were made by ClustalW or MUSCLE. Phylogenetic tree was built using *phylogeny.fr (Dereeper et al., 2010)* with default parameters, following the MUSCLE alignment. One hundred replicas were used in the bootstrap test. Viral sequences used to build the tree are listed in Table II.3. The Protean computer program (part of the DNASTAR Lasergene8 software suite) was used for isolelectric point (pI) calculations. Protein quantifications in western blot and *in vitro* translation assays were performed with Quantity one or Image J software.

Table II.3. Viral species and GenBank accession numbers for viral protein sequences used in the study.

Acronym	Species	GenBank	Genus
PPV	*Plum pox virus*	NP734339	*Potyvirus*
CVYV	*Cucumber vein yellowing virus*	YP308878	*Ipomovirus*
TuMV	*Turnip mosaic virus*	NP734213	*Potyvirus*
TVMV	*Tobacco vein mottling virus*	NP734336	*Potyvirus*
AgMV	*Agropyron mosaic virus*	YP054392	*Rymovirus*
BstMV	*Brome streak mosaic virus*	NP734253	*Tritimovirus*

WSMV	*Wheat streak mosaic virus*	AAM48213	*Tritimovirus*
BlVY	*Blackberry virus Y*	YP851199	*Brambyvirus*
UCBSV	*Ugandan cassava brown streak virus*	CBA13048	*Ipomovirus*
TriMV	*Triticum mosaic virus*	*ACT53745*	*Poacevirus*

II.11 Preparation of BYL extract

Five milliliters of 7-day-cultured BY-2 cell suspension were diluted with 45 ml fresh MS medium in a 250 ml Erlenmeyer flask and cultured for 4-5 days at 26-27 ºC, 160 rpm and darkness. Ca. 45 ml of this cell suspension were transferred to a 50-ml conical tube. Cells were collected by centrifugation at 100 *g* for 1 min at room temperature (no brake) obtaining a packed cell volume of 8-15 ml. Twenty five ml of protoplast washing buffer [0.37 mM mannitol, 5 mM $CaCl_2$, 12.5 mM sodium acetate (pH 5.8)] with lytic enzymes [1% cellulase Onozuka RS (Yakult Pharmaceutical Industry) and 0.2% Pectolyase Y23 (Kyowa Chemical Products)], previously filtered through a 0.45 µl filter membrane, were added to the cell pellet, and the resultant mixture was transfered to a 250 ml Erlenmeyer flask. Cells were incubated at 26-30 ºC with gentle swirling (30 rpm) for 1.5-2 hours (Murota et al., 2011).

Approximately 1 ml of the treated cells was mixed with 2 ml of 5% Percoll solution. This mixture was overlaid on a 6-ml Percoll gradient 10–35% (v/v) layered on top of 3 ml of 55% (v/v) Percoll solution set up in an open-top centrifuge tube (#361707, Beckman Coulter). All Percoll solutions contained 0.7 M mannitol, 20 mM $MgCl_2$ and 5 mM PIPES-KOH (pH 7.4). Gradient centrifugation was performed at 12,000 g for 90 min at 25 ºC in a SW41 rotor (Beckman Coulter) with slow acceleration and no braking. After centrifugation, protoplasts, concentrated in the interface of the 5%-10% Percoll layers, and evacuolated protoplasts, concentrated in the interface of the 35%-55% Percoll layers, were collected. Three volumes of ice-cold 0.7 M mannitol were added to the samples and centrifuged at 100 x g for 5 min at room temperature (no break). Washing process was repeated three times before final resuspension in 4 volumes of ice-cold 0.45-µm membrane-filtered TR buffer [30 mM HEPES-KOH, pH 7.4, 80 mM potassium acetate (KAc), 1.8 mM magnesium acetate, 2 mM dithiothreitol, one tablet per 50 ml of Complete, EDTA-free, protease inhibitor cocktail (Roche)]. Cells were disrupted on ice with a Dounce homogenizer (approximately 300 strokes) and nuclei and non-disrupted cells were pelleted by centrifugation for 10 min at 500 g at 4 ºC (Murota et al., 2011).

The collected supernatants (~2.5 ml of EP and 5 ml P) were gel-filtrated using a pre-packed PD-10 column (GE Healthcare Life Science) and TR buffer to remove endogenous low molecular weight factors and other possible contaminants. BYL fractions were tested for their ability to complement PPV P1 cleavage activity in RRL *in vitro* translation system and stored at -80 ºC (Murota et al., 2011).

II.12 Ion exchange chromatography

In a first experiment, anion exchange chromatography was conducted in a 1-ml Q resin column (GE Healthcare Life Sciences) coupled to a fast protein liquid chromatography (FPLC) device. Column was equilibrated with pre-cold TR buffer with 200 mM KAc (no protease inhibitors were included in this or any of following purification steps). BYL was defrosted on ice and centrifuged at 7,200 g for 5 min at 4 ºC. Supernatant was collected and KAc concentration was adjusted to 200 mM before loading into the column. Column was washed until absorbance reached a flat line and step-elution was performed with pre-cold TR buffer including increasing concentrations of KAc (300 mM, 350 mM, 400 mM, 450 mM, 500 mM, 550 mM, 600 mM and 1 M). Flow-through (FT) and 8 fractions of 1 ml were collected for each elution step, concentrated up to ~80 µl by centrifugation in Amicon Ultra 0.5 ml filter devices MWCO 3 kDa (Sigma) and stored at – 80 ºC.

In following experiments, anion exchange chromatography was performed in a home-made column of ~300 µl Q resin (Q Sepharose High Performance, GE Healthcare Life Sciences) mounted in a 1-ml syringe. In this case, the elution steps were conducted with pre-cold TR buffer including 300 mM and 500 mM KAc. FT and 2 fractions of 1 ml were collected for each elution step, concentrated up to ~80 µl as explained above, and stored at -80 ºC.

Cation exchange chromatography was conducted in home-made columns similar to those described above, containing ~300 µl of F resin (SP Sepharose Fast Flow, GE Healthcare Life Sciences). F column was equilibrated with pre-cold TR buffer with 50 mM KAc. Concentrated fractions eluted (EF) from the Q resin column by TR buffer (500 mM KAc) were diluted 10 times with pre-cold TR buffer (without KAc) to reduce salt concentration to 50 mM before loading into the equilibrated F column. Column was washed with pre-cold TR buffer (50 mM KAc) and a step-elution was performed with pre-cold TR buffer with increasing KAc concentration up to 1 M (including 100 mM, 200 mM, 300 mM, 400 mM, 500 mM and 1 M). Flow-through and 6 fractions of 1 ml were collected for each elution step, concentrated up to

~80 µl by centrifugation in Amicon Ultra 0.5 ml filter devices MWCO 3 kDa (Sigma) and stored at – 80 ºC.

II.13 Gel filtration chromatography

Samples purified by the two steps of ion exchange chromatography (~200 µl) were loaded in a column of Superdex 200 10/300 GL (GE Healthcare Life Science) previously equilibrated with pre-cold TR buffer (80 mM KAc) and connected to an FPLC system. Gel filtration was performed in this same buffer. Fractions of 500 µl were collected and concentrated up to ~50 µl as described above before being tested in the RRL P1 protease complementation assays. Selected fractions were used for mass spectrometry analysis.

II.14. Mass spectrometry

II.14.1 In-gel protein digestion and MALDI peptide mass fingerprinting

Bands of interest from coomasie-stained gels of the first BYL purification experiment were excised manually, deposited in 96-well plates and processed automatically in a Proteineer DP (Bruker Daltonics, Bremen, Germany). The digestion protocol used was based on Shevchenko *et al.* (1996) with minor variations: gel plugs were washed firstly with 50 mM ammonium bicarbonate and secondly with acetonitrile (ACN) prior to reduction with 10 mM dithiothreitol (DTT) in 25 mM ammonium bicarbonate solution, and alkylation was carried out with 55 mM indole-3-acetic acid (IAA) in 50 mM ammonium bicarbonate solution. Gel pieces were then rinsed firstly with 50 mM ammonium bicarbonate and secondly with ACN, and then were dried under a stream of nitrogen. Proteomics Grade Trypsin (Sigma Aldrich) at a final concentration of 16 ng/µl in 25% ACN/50 mM ammonium bicarbonate solution was added and the digestion took place at 37 ºC for 4 h. The reaction was stopped by adding 50% ACN/0.5% Trifluoroacetic acid (TFA) for peptide extraction. The eluted tryptic peptides were dried by speed-vacuum centrifugation and were resuspended in 4 µl of MALDI solution (30% ACN/15% isopropanol/0.5% TFA). A 0.8 µl aliquot of each peptide mixture was deposited onto a 384-well OptiTOF™ Plate (AB SCIEX, Foster City, CA) and allowed to dry at room temperature. A 0.8 µl aliquot of matrix solution [3 mg/mL α-Cyano-4-hydroxycinnamic acid (CHCA) in MALDI solution] was then deposited onto dried digest and allowed to dry at room temperature.

For Matrix Assisted Laser Desorption Ionization-Time of Flight/Time of Flight (MALDI-TOF/TOF) analysis, samples were automatically acquired in an ABi 4800 MALDI TOF/TOF mass spectrometer (AB SCIEX, Foster City, CA) in positive ion reflector mode (the ion

acceleration voltage was 25 kV to MS acquisition and 2 kV to MSMS) and the obtained spectra were stored into the ABi 4000 Series Explorer Spot Set Manager. Peptide mass fingerprint (PMF) and MSMS fragment ion spectra were smoothed and corrected to zero baseline using routines embedded in ABi 4000 Series Explorer Software v3.6. Each PMF spectrum was internally calibrated with the mass signals of trypsin autolysis ions to reach a typical mass measurement accuracy of <25 ppm. Known trypsin and keratin mass signals, as well as potential sodium and potassium adducts (+21 Da and +39 Da) were removed from the peak list. To submit the combined PMF and MS/MS data to MASCOT software v.2.3.02 (Matrix Science, London, UK), GPS Explorer v4.9 was used, searching in a customized *N. tabacum* protein database downloaded from Sol Genomics Network (https://solgenomics.net/) containing 85439 sequences. The following search parameters were used: enzyme, trypsin; allowed missed cleavages, 1; carbamidomethyl cystein as fixed modification by the treatment with iodoacetamide; variable modifications, oxidation of methionine; mass tolerance for precursors was set to ± 50 ppm and for MS/MS fragment ions to ± 0.3 Da. The confidence interval for protein identification was set to ≥ 95% ($p < 0.05$) and only peptides with an individual ion score above the identity threshold were considered correctly identified. After the first Mascot search, peptides corresponding to the first hit were removed from the list and a second search was launched with the remaining peptides.

II.14.2 Liquid chromatography and mass spectrometry analysis

Samples from gel filtration chromatography were treated individually in 20 µl 7 M Urea/2 M Thiourea/100 mM triethylammonium bicarbonate (TEAB), pH 7.5, reduced with 2 µL of 50 mM Tris (2-carboxyethyl) phosphine (TCEP, SCIEX), pH 8.0, at 37°C for 60 min and followed by addition of 1 µL of 200 mM cysteine-blocking reagent (methyl methanethiosulfonate, Pierce) and incubation for 10 min at room temperature. Urea concentration was reduced to 2 M with 25 mM TEAB and samples were digested with trypsin overnight at 37º C with a substrate: enzyme ratio of 20:1. Digested samples were desalted with Stage-Tip C18 (Thermo Fisher Scientific, MA, USA) and dried in a speed-vac centrifuge. Subsequently, Nano Liquid Chromatography coupled to Electrospray Tandem Mass Spectrometry (LC ESI-MS/MS) analysis was performed using an Eksigent 1D- nanoHPLC coupled to a 5600 TripleTOF QTOF mass spectrometer (SCIEX, CA, USA). Gradient elution was performed according to the following scheme using A (2% ACN, 0.1% formic acid) and B (100% ACN, 0.1% formic acid) solutions: isocratic conditions of 95% A: 5% B for one minute, a linear increase to 50% B in 15 min, a linear increase to 90% B in 30 seconds, isocratic conditions of 90% B for five minutes and return to

initial conditions in 30 seconds. Automatic data-dependent acquisition using dynamic exclusion allowed to obtain both full scan (m/z 350-1250, 250 msec) MS spectra followed by tandem MS CID (Collision-induced dissociation) spectra (100 msec) of the 30 most abundant ions per MS spectrum.

MS and MS/MS data were used to search against a customized target *N. tabacum* protein database downloaded from Sol Genomics Network (https://solgenomics.net/) containing 85439 sequences. Database searches were done using a licensed version of Mascot v.2.6.1. Peptide mass tolerance was set at 25 ppm and 0.05 Da for MS and MS/MS spectra, respectively, and 2 missed cleavages were allowed. Mascot score threshold for peptide identification was set to a value equal or higher than 20.

II.15 P1 protease activity complementation assays

RRL *in vitro* translation system (Promega) was used for the P1 protease activity complementation assays. mRNA preparation and *in vitro* translation in RRL were performed as detailed in section II.7. To maintain constant the 5 µl volume of the reaction, the standard amount of RRL suggested by the manufacturer (3.5 µl) was reduced to 1.4 µl and complemented with 2.1 µl of the corresponding supplement unless otherwise specified.

For the inhibition assay, TUDOR-SN 1 nuclease inhibitor, deoxythymidine 3′, 5′-bisphosphate (pdTp), was resuspended in water to a final concentration of 50 mM. From this stock, 1 mM, 5 mM and 10 mM pdTp aliquots were prepared. mRNA preparation and *in vitro* translation in RRL were performed as detailed in section II.7. To maintain constant the 5 µl volume of the reaction, the standard amount of RRL suggested by the manufacturer (3.5 µl) was reduced to 1.85 µl and complemented with 1.4 µl BYL extract and 0.25 µl of H_2O or 1 mM, 5 mM and 10 mM pdTp.

II.16 P1 processing activity in *N. benthamiana* plants in which candidate genes have been down-regulated by virus-induced gene silencing (VIGS)

TRV was used as vector for VIGS by cloning about 350-450-bp fragments of each candidate gene in the plasmid carrying the viral sequence of its second RNA, pTRV2 (Senthil-Kumar & Mysore, 2014). Sequences to be cloned were selected according to the VIGS tool of Sol Genomics Network (https://solgenomics.net/). In some cases, several gene fragments were cloned in independent pTRV2 plasmids to warrant efficient silencing of the candidate genes. Cloning was performed as detailed in section II.5.1.

Agrobacterium strains carrying pTRV1 plasmid (OD_{600}=0.5) and one or several modified pTRV2 plasmids (OD_{600}=0.5, in total) were co-agroinoculated in the two youngest leaves of 3 weeks-old *N. benthamiana* plants, as indicated in section II.3. Twelve days later (7 days in the case of TRV2-HSP70), pP1HCTAP and pTAPHC transient expression plasmids were agroinfiltrated in the new 2 youngest leaves. Tissue was collected after two days and analyzed by western blot using directly the secondary antibody, which recognizes the TAP tag.

RESULTS

III. Results

III.1. P1 protein classification in the *Potyviridae* family

P1 leader proteases are the most divergent potyviral proteins in length and amino acid sequence (Adams et al., 2005b;Valli et al., 2007). Despite this large variability, the P1 C-terminal region is relatively well conserved. It harbors a serine protease domain that is responsible for *cis*-cleavage of the P1-HC junction and thus, P1 self-releases from the remainder of the polyprotein (Adams et al., 2005b). Once released, the mature P1 C-terminal end is thought to be trapped in the active cleft, leading to self-inhibition of the *trans* cleavage activity (Verchot et al., 1992). The study of mechanistic differences of the two P1 proteins of CVYV (Rodamilans et al., 2013) suggested that all P1 proteins in the family could be classified in two groups based on the requirement for a plant factor(s) for protease activity and other features, like shared isoelectric point or RSS activity. The following section explores these features in different viruses and genera of the *Potyviridae* family as well as studies the *in-trans* proteolytic activity of these proteins.

III.1.1 Self-cleavage activity in WGE and RRL of potyvirid P1 proteins

According to previous works (Carrington *et al.*, 1990;Verchot et al., 1992), Type A P1 proteins are able to perform autocatalytic activity in WGE, but not in RRL. On the other hand, Type B P1 proteins can work in both translation systems with high efficiency (Rodamilans et al., 2013). In order to classify potyvirid P1 proteins as Type A or B, based on their ability to hydrolyze in the presence or absence of plant factors, a series of translation experiments were conducted using RRL or WGE translation systems. P1 orthologs from the potyviruses PPV, TuMV and TVMV, ipomoviruses CVYV and UCBSV, the tritimovirus BStMV, and the brambyvirus BlVY were tested. Whereas all potyviral orthologs, as well as the ipomoviral CVYV P1a, only cut in the WGE system, bands with the expected mobility of self-cleaved P1 proteins of the ipomovirus UCBSV, the tritimovirus BStMV and the brambyvirus BlVY were detected in both *in vitro* translation systems (Fig. III.1). These products originated from the P1 protease activity as they were not present in non-functional mutants (S355A for BStMV P1, S692A for BlVY P1 and S308A for UCBSV P1).

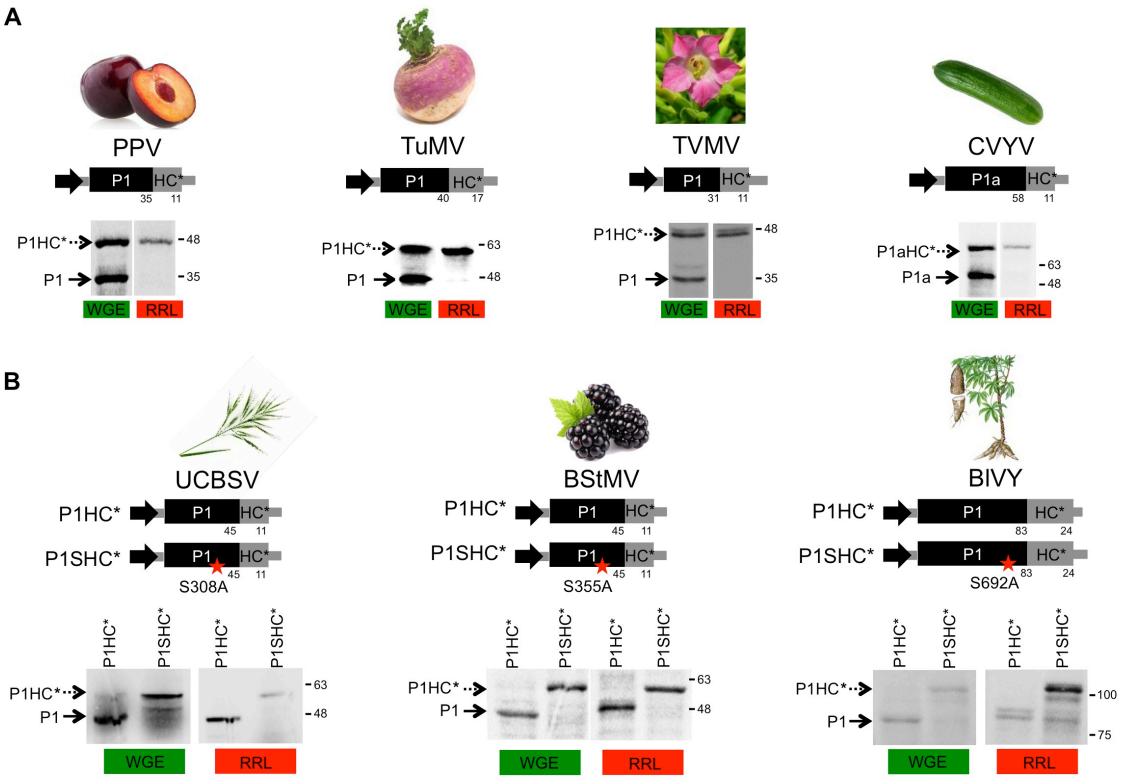

Figure III.1. Proteolytic activity of P1 proteins from four monopartite genera of the family *Potyviridae*. Self-cleavage activity was evaluated in RRL and WGE *in vitro* translation mixtures. **(A and B)** Diagrams of DNA constructs of the coding sequences of leader proteases from PPV, TuMV, TVMV and CVYV (panel **A**), and from UCBSV, BStMV and BlVY (panel **B**), amplified by PCR and subjected to *in vitro* transcription and further *in vitro* translation, are shown on top of the panels. Mutations in the serine residue of the active center of the protein are indicated with a red asterisk. T7 RNA polymerase promoter (black arrow) drives the PPV 5'UTR and the coding sequence of the protease (black box) and 97 N-terminal amino acids from HC (HC*, grey box) followed by a stop codon and PPV 3'UTR. Molecular weight of the expected protein products is also shown. On the bottom part of the panels, ^{35}S-labeled products were resolved by SDS-PAGE and detected by autoradiography. Processed and unprocessed products are marked (left); right, molecular weight markers.

III.1.2 RNA silencing suppression activity of potyvirid P1 proteins

Plant factor(s) requirement is not the only feature used to classify P1 proteins as Type A or B. The ability to suppress RNA silencing has been postulated as another characteristic defining P1 proteins. Thus, in the case of the ipomovirus CVYV, P1b presented RSS activity while P1a did not. This function was independent of the protease activity. Same is true for potyviral PPV P1, a Type A P1 protein, which does not bear RSS activity *per se*, although it has been reported to act as an enhancer of the RSS activity of HC (Valli et al., 2008;Valli et al., 2006). To further characterize leader proteins in the *Potyviridae*, P1 and HC proteins of the tritimovirus BStMV and the brambyvirus BlVY were selected. The former one is anticipated to bear RSS activity since P1s of other members of the same genus, such as WSMV, has it (Tatineni *et al.*, 2010). For

the latter one there is no previous data available. Transient silencing assays were performed in *N. benthamiana* plants using GFP expression as trigger of silencing and reporter as described (Voinnet et al., 2003). Delivery of BStMV P1 allowed maintaining high levels of GFP at 6 days post-agroinfiltration (dpa), showing that this protein has RSS activity. Co-expression of GFP with BStMV HC rendered similar levels of fluorescence to the co-expression with an empty vector indicating no suppression of RNA silencing in this case (Fig. III.2A and C). Neither BlVY P1 or HC proteins showed increased levels of GFP fluorescence compared to co-expression with an empty vector, indicating that RNA silencing was not suppressed. In agreement with previous reports (Carbonell et al., 2012), CVYV P1b alleviated the host RNA silencing activation and allowed high levels of GFP expression at 6 dpa (Fig. III.2B and D).

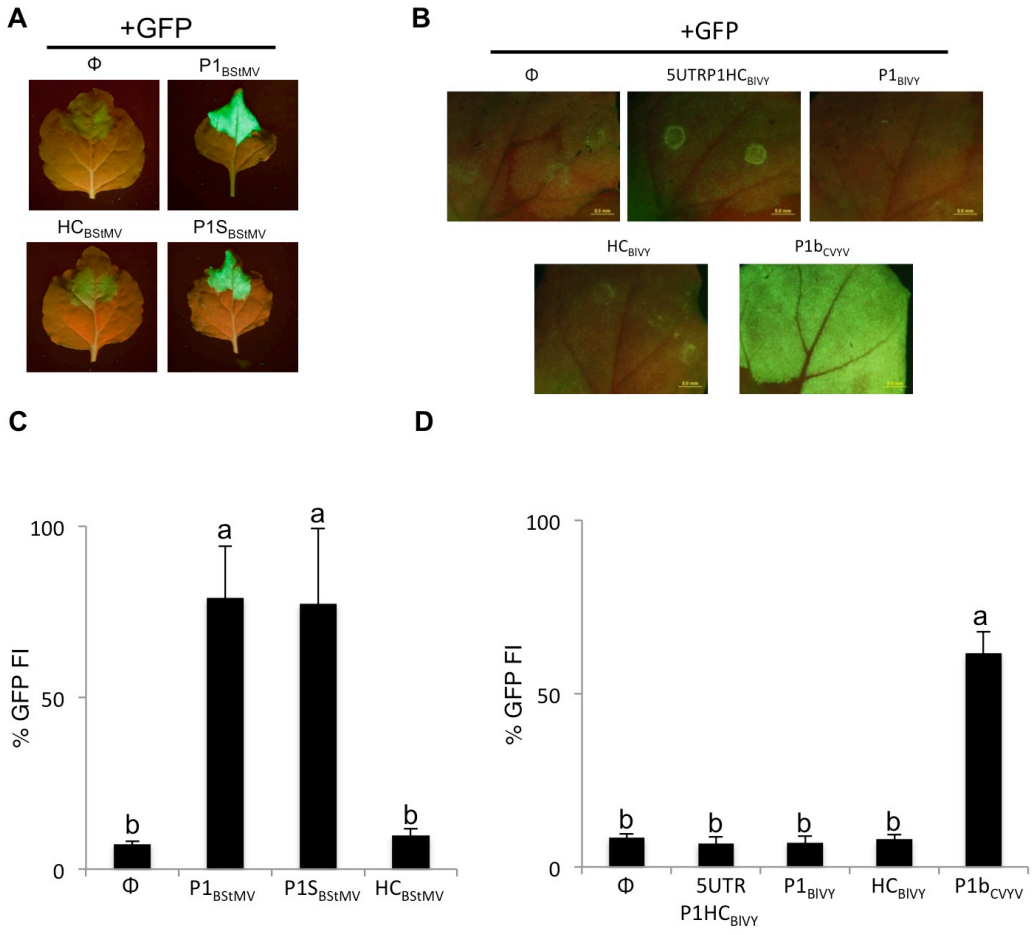

Figure III.2. RSS activity of BStMV and BlVY leader proteinases. (A and B) *N. benthamiana* leaves were co-agroinfiltrated with a GFP-expressing construct plus constructs expressing the indicated BStMV (A) or BlVY (B) proteins; CVYV P1b was used as positive control in (B) and an empty vector (Φ) was used as negative control in both cases. Pictures were taken at 6 dpa. **(C)** GFP fluorescence intensity (FI) of agroinfiltrated leaves was quantified in a 96-well plate reader at 6 dpa. Bars marked with two asterisks correspond to samples with a statistically significant difference ($p<0.01$) when compared to the negative control by one-way Anova and Tukey's HSD test.

III.1.3 P1 serine proteases do not display *in trans* activity

P1 leader proteases have been previously described as enzymes acting *in cis* and co-translationally (Valli et al., 2006;Verchot & Carrington, 1995b). These studies were conducted with TEV P1 and CVYV P1a. The inability to perform *in trans* has been related to a blockage of the active cleft by the C-terminal amino acid of the protease upon release from the rest of the polyprotein (Verchot et al., 1992). Similar behavior was described for the TuMV HC protease domain (Guo *et al.*, 2011), and this same proteolytic pattern can be observed for togavirin, another viral serine protease that also acts *in cis* presenting the C-terminal amino acid embedded in the catalaytic pocket. As mentioned in the introduction, togavirin, the leader protease of alphaviruses, is a versatile protease that is able to bind RNA and assemble into the viral capsid (Krupovic & Koonin, 2017). Interestingly, upon deletion of the last amino acid, Trp267, togavirin changes its mode of action and perform *in trans* (Aggarwal *et al.*, 2014). Considering this result, it was hypothesized that leader serine proteases of the *Potyviridae* family might display a similar behavior. The following experiments were intended to clarify this issue and verify a possible *in cis/in trans* switching mode upon deletion of C-terminal amino acids.

PPV P1 and CVYV P1b were tested. Constructs of P1SHC* and P1bSTAP with their respective catalytic serines mutated (S259A for PPV P1 and S264A for CVYV P1b) were used as *in trans* targets for the proteases. An N-terminal-deleted version of PPV P1 ($P1Pro_{C0}$) and the full-length CVYV P1b ($P1b_{C0}$) reported to work *in cis* were used as negative controls (Pasin et al., 2014b;Rodamilans et al., 2013). C-terminal deletions of 1, 3 and 5 amino acids of these proteases were tested ($P1Pro_{C1,3,5}$ and $P1b_{C1,2,3}$) (Figure III.3A). Experiment was carried out using WGE under standard conditions for protein translation. Figure III.3B shows that bands corresponding to the size of the unprocessed products can be observed as well as bands corresponding to P1Pro and P1b and the corresponding C-terminal deletion mutants. No bands, however, were observed of the anticipated size of P1S or TAP indicating that P1pro and P1b are not changing their activity from an *in cis* to an *in trans* mode (Figure IIIB and C) upon deletion of C-terminal amino acids.

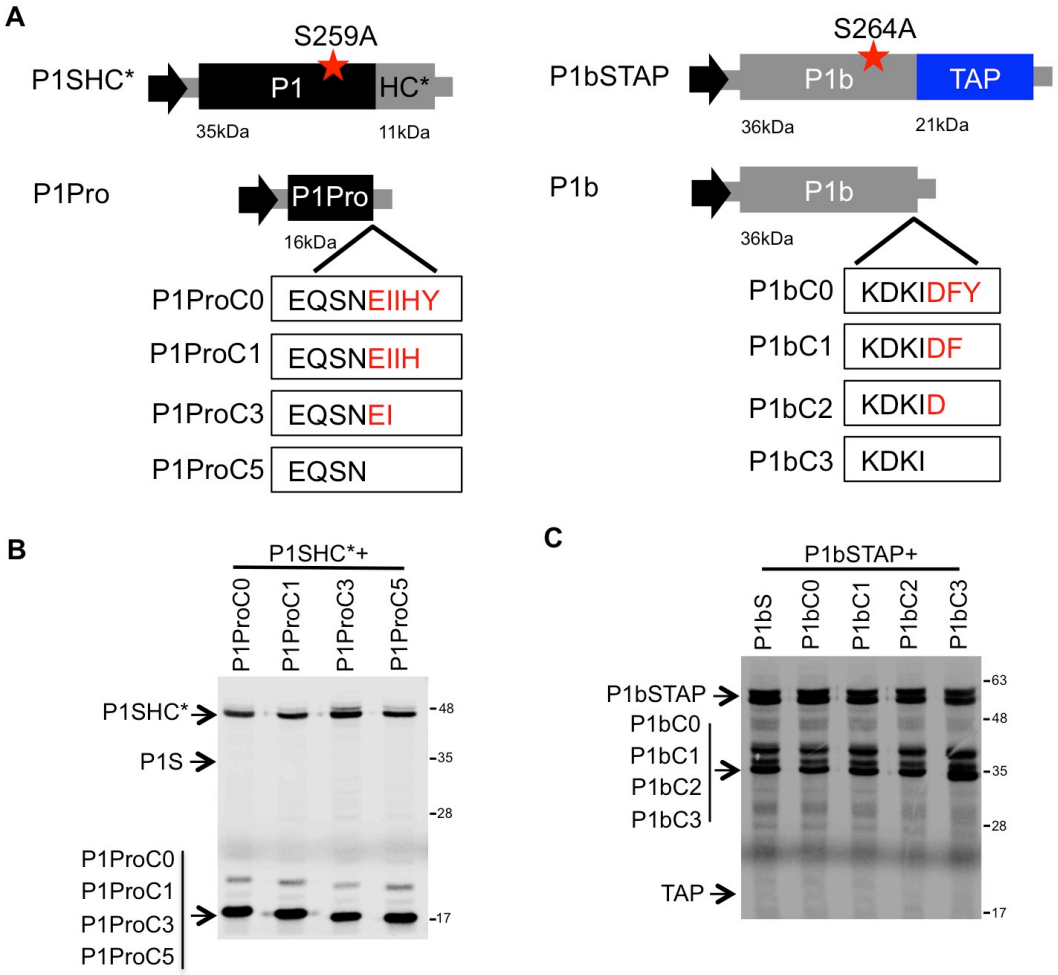

Figure III.3. PPV P1 *in-trans* proteolytic activity assay in WGE. (A) Diagrams corresponding to the different PPV P1 and CVYV P1b constructs used. Mutations in the serine residue of the active center in P1SHC* and P1bSTAP used as substrate for cleavage are marked with a red star. C-terminal deletion mutants are indicated with a C followed by the number of amino acids removed (0, 1, 3 or 5 for P1S, 0, 1, 2 or 3 for P1b). T7 RNA polymerase promoter drives the PPV 5'UTR and the coding sequences of PPV P1 (black box) and 97 N-terminal amino acid of PPV HC (HC*, grey box) or the coding sequences of P1b (black box) and TAP (blue box). HC* and TAP are followed by a stop codon and the PPV 3'UTR. Expected translation products and their molecular weights are displayed above. **(B and C)** *trans*-proteolytic activity assay of PPV P1Pro (B) and CVYV P1b (C) in WGE *in vitro* translation mixture. ^{35}S-labeled products were resolved by SDS-PAGE and detected by autoradiography. Individual proteins and unprocessed products are marked with arrows on the left side of the autoradiograph; right, molecular weight markers.

To further explore this result, the effect of co-expression of P1SHC* and P1Pro$_{C0,1,3,5}$ deletion mutants was analyzed in a time course of 2, 5 and 16 h in WGE. As happened in the previous experiment, bands corresponding to the size of the expressed products could be observed, but there was no band corresponding to the processed P1S (Figure III.4).

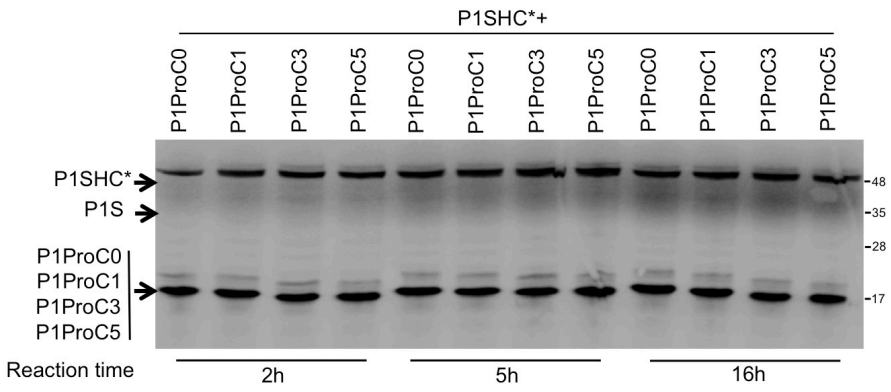

Figure III.4. Time course of *trans*-proteolytic activity of PPV P1pro in WGE. *In vitro* translation samples were collected at 2 h, 5 h and 16 h. ^{35}S-labeled products were resolved by SDS-PAGE and detected by autoradiography. Individual proteins and unprocessed products are marked with arrows on the left side of the autoradiograph; right, molecular weight markers.

III.2. Host range definition of potyviral P1 proteins

Part of the results presented in the following section have been reported in our own published articles (Shan *et al.*, 2018;Shan *et al.*, 2015).

Potyvirid Type A P1 proteins have been involved in host range definition (Rodamilans et al., 2013). Previous work by Carbonell, et al (2012) suggests that the inability of a chimeric PPV, carrying CVYV P1a-P1b instead of P1-HC, to infect *N. benthamiana* plants is due to a poor proteolytic activity of P1a. Following this hypothesis several experiments were conducted in two hosts, *N. benthamiana* and *C. sativus*, using PPV as a model varying P1 proteases in order to modify host specificities. Further studies with transient expression clones help to clarify viral infection results.

III.2.1. CVYV P1a and PPV P1 in *Nicotiana benthamiana*

III.2.1.1 Viral accumulation of chimeric PPV bearing P1a is increased in the presence of an extra NIa cleavage site between P1a and HC in *N. benthamiana*

A series of chimeric viruses were prepared using PPV as backbone. P1 was replaced by P1a (PPV-P1a) or by P1a followed by a cleavage site of the PPV proteinase NIa, thus placed in front of HC (PPV-P1aNIa). The extra cleavage site by NIa will ensure release of P1a from the rest of the polyprotein and should fix any host compatibility problems that PPV-P1a might be carrying as consequences of a host-specific defect of the proteolytic activity of CVYV P1a. PPV-P1aS,

with a mutation in the catalytic serine of P1a, was used as negative control. PPV-P1aP1b chimeric virus (Carbonell et al., 2012) was also used as control (Figure III.5A).

All cDNA constructs were delivered to *N. benthamiana* plants by agro-infiltration and virus spread was monitored using GFP fluorescence in inoculated and upper tissue, at 10 and 14 dpa, respectively (Figure III.5B). Viral accumulation was measured by anti-CP western blot analysis; release of mature HC was detected by anti-HC immunoblot assay (Figure III.5C, D). Plants treated with the PPV cDNA clone carrying the P1a protease mutant (PPV-P1aS) did not show any sign of local or systemic infection (PPV-P1aS; Figure III.5). Replacement of P1 sequence by wild-type CVYV P1a protein (PPV-P1a) severely impaired viral accumulation and only a faint CP signal was detected in western blot analysis of upper leaves (Figure III.5D). Inclusion of an NIa cleavage site between wild-type P1a and HC significantly enhanced viral accumulation ($p < 0.05$, PPV-P1aNIa versus PPV-P1a), which reached levels close to the wild-type PPV clone in both inoculated and upper leaves (Figure III.5C, D). As previously reported (Carbonell et al., 2012), the PPV-P1aP1b chimera presented severe defects in viral infection.

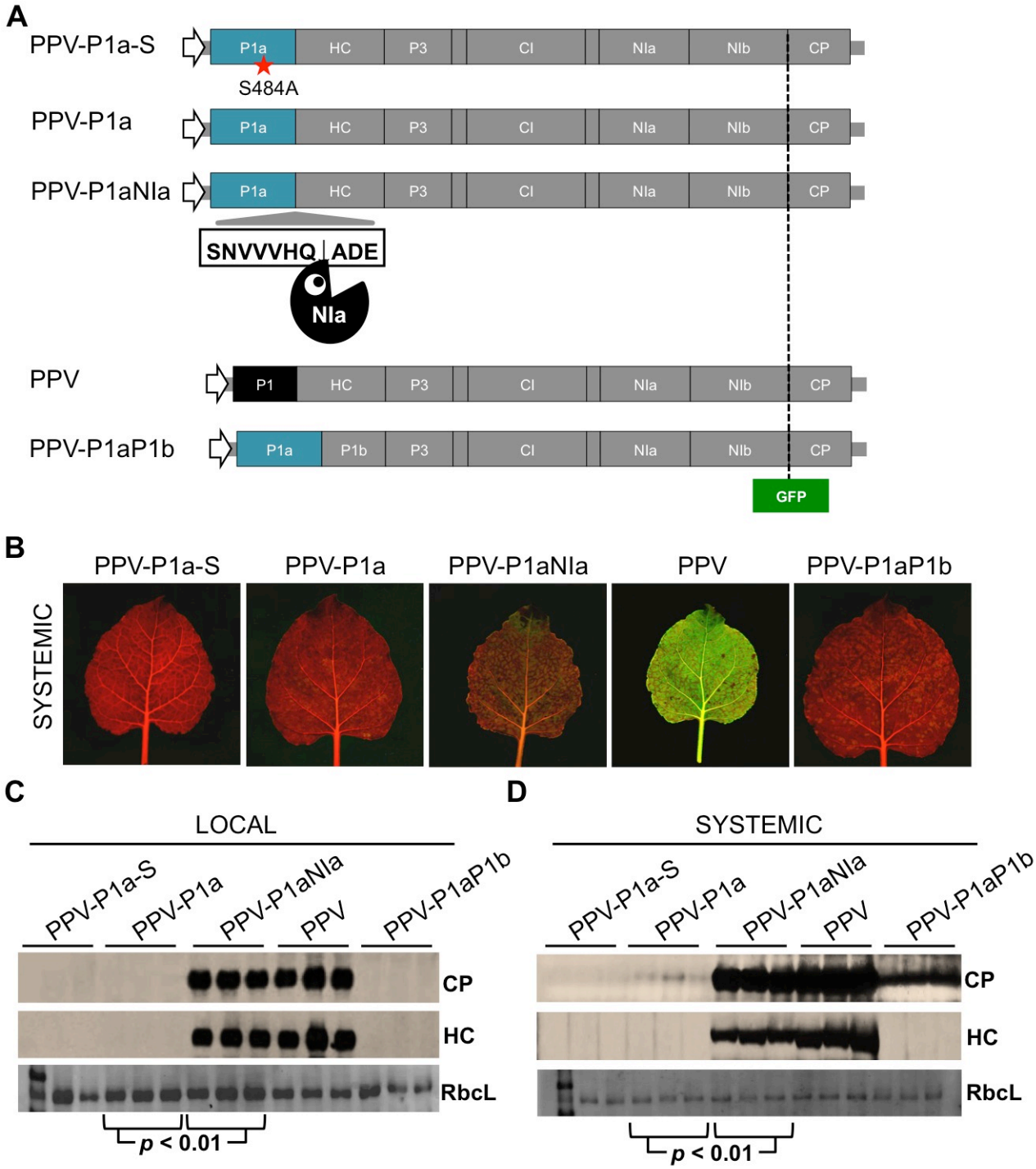

Figure III.5. Viral infection assay in *N. benthamiana* plants. (**A**) Schematic representation of the chimeric viruses. (**B**) PPV-derived chimeras were delivered to *N. benthamiana* plants by agro-inoculation and samples analyzed at 10 dpa (LOCAL, infiltrated leaves) and 14 dpa (SYSTEMIC, upper non-inoculated leaves). GFP photos were taken on a blue light transilluminator. (**C**, **D**) Anti-CP and anti-HC western blot analyses; the differences in CP signal intensity between the indicated samples are statistically significant by Student *t*-test ($p < 0.01$). Each lane corresponds to a sample pool from one/two agroinfiltrated plants. Ponceau red-stained blots showing the RubisCO large subunit (RbcL) were included as loading controls.

III.2.1.2 CVYV P1a shows incomplete catalytic activity and the p19 silencing suppressor complements P1a defects in *N. benthamiana*

Viral infection assay in *N. benthamiana* showed that introduction of an extra NIa cleavage site between P1a and HC rescued PPV infectivity. This result strongly suggests that the defects observed in PPV chimeras were linked to the inability of P1a to correctly perform autocleavage in this host. To verify processing of P1a-HC precursor in *N. benthamiana* plants, a transient expression assay was done in which series of constructs were co-agroinfiltrated with the strong RNA silencing suppressor p19 from *Tomato bushy stunt virus* (TBSV) (Voinnet et al., 2003) to provide an even silencing suppression effect (Figure III.6A). The protease domain of the NIapro and GFP were also co-expressed with all samples to reproduce previous experimental conditions and to verify the proper silencing suppression (Figure III.6B).

Anti-HC polyclonal antibody was used in western blot analysis to detect mature HC as well as uncleaved protein precursors (Figure III.6C). P1S, with a mutation in the catalytic serine of P1, was included as background reference. Samples expressing this construct showed a mayor band corresponding to a protein of ~87 kDa, close to the expected size of the P1-HC fusion product. In the case of P1 and P1aNIa constructs, a strong band matching the expected size of free HC (~52 kDa) was detected with no appreciable bands fitting the polyprotein precursors. In the case of the P1a construct that lacks the NIa cleavage site, both bands corresponding to HC (~52 kDa) and the P1a-HC unprocessed product (~113 kDa) could be observed (Figure III.7C). This result indicates that in *N. benthamiana* both PPV P1 and NIa proteases outperform CVYV P1a cleavage capacities, in agreement with previous reports (Carbonell et al., 2012).

Like P1 mutants deficient in protease activity (Pasin et al., 2014b), incomplete self-cleavage of P1a in *N. benthamiana* might affect activity of a downstream RSS and limit infectivity of P1a-containing viral clones. Supply of p19 RSS allowed propagation of suppressor-deficient TuMV clone in a transient infection assay (Garcia-Ruiz *et al.*, 2010). Thus, PPV-P1aS (Figure III.5A) was co-expressed with p19 or empty vector (Φ). Consistent with previous results, PPV-P1aS clone was non-infectious, as shown by absence of GFP fluorescence and CP accumulation in leaves co-infiltrated with the empty vector (Figure 6D, E). However, p19 co-expression rescued PPV-P1aS deficient clone. GFP fluorescence intensity and PPV CP accumulation levels were significantly higher in the presence of p19 compared to the empty vector control (Figure III.6D, E).

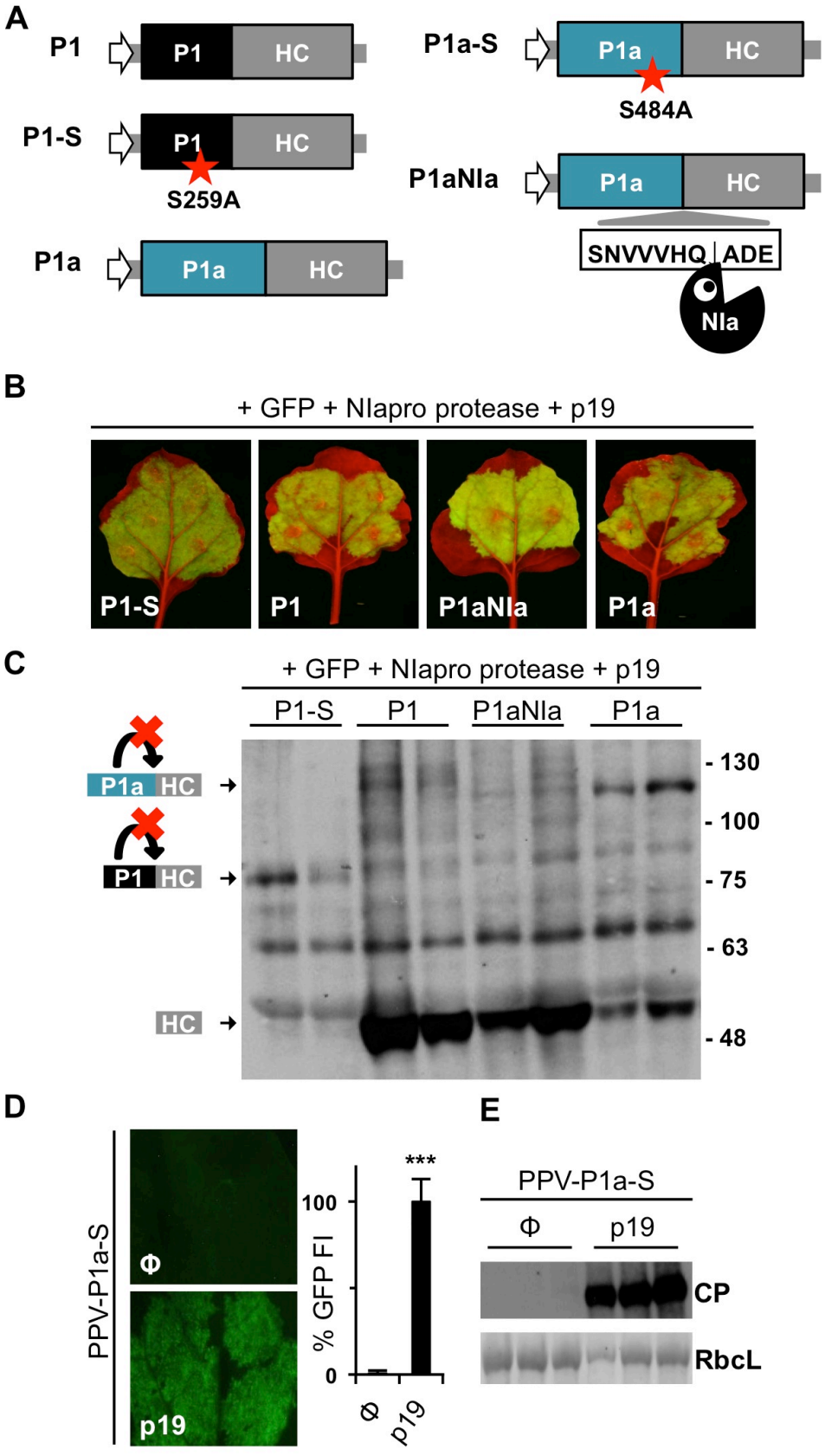

Figure III.6. Processing of different fusion constructs of HC and requirement of RSS activity in a transient infection assay. (A) Schematic representation of the transient expression constructs used. (**B and C**) *N. benthamiana* plants were co-agroinfiltrated with plasmids for the expression of GFP, NIa protease, p19 RSS plus

the corresponding experimental plasmids, and analyzed at 6 dpa. (**B**) Pictures of agroinfiltrated leaves taken on a blue light transilluminator. (**C**) Anti-HC western blot analysis of the agroinfiltrated tissue. Processed and unprocessed products are marked with arrows on the left side of the film; right, molecular weight markers. Each lane corresponds to a sample pool from two agroinfiltrated plants. (**D and E**) *N. benthamiana* plants were co-agroinfiltrated with PPV-P1aS and either p19 or empty vector (Φ). (**D**) Pictures, taken under an epifluorescence microscope at 6 dpa, of locally infect leaves. GFP fluorescent intensity (FI) was quantified in a 96-well plate reader. Relative GFP signal intensities are indicated using p19 mean value equal to 100; the difference between the values is statistically significant by Student *t*-test ($p < 0.001$, $n = 4$ biological replicates). (**E**) Viral accumulation was assessed by anti-PPV CP western blot. Each lane corresponds to a single agroinfiltrated plant. Ponceau red-stained blot showing RbcL was included as loading control.

III.2.1.3 *cis*-supply of CVYV P1a impairs HC silencing suppression activity, which is restored by NIa protease-mediated cleavage in *N. benthamiana*

Previous assays show that CVYV P1a presents restricted autoproteolytic ability in *N. benthamiana* and this deficiency encumbers the infective capacity of PPV chimeras carrying the CVYV P1a coding sequence. This defect appears to be a consequence of a lack of RSS ability on the part of HC (Figures III.5 and III.6). To verify this, transient expression constructs previously designed were used to co-agroinfiltrate *N. benthamiana* plants with and without a plasmid encoding PPV NIapro *trans*-cleaving protease to modulate cleavage efficiency (Figure III.7A).

Constructs were delivered to *N. benthamiana* and GFP accumulation was measured at 6 dpa (Figure III.7). As anticipated, independently of the presence of NIapro protease, P1 construct released a functional HC silencing suppressor and allowed high levels of GFP accumulation. Construct P1a, on the other hand, failed to sustain reporter accumulation and behaved as the negative control (P1aS) in both situations (Figure III.7). Inclusion of an NIa cleavage site between wild-type P1a and HC (P1aNIa) significantly increased GFP fluorescence intensity ($p < 0.01$) to levels similar to the positive control. This only happened in samples co-agroinfiltrated with NIapro protease. No significant differences were detected between P1a constructs when the NIapro protease-expressing strain was not included in agroinfiltration mixes; in this case, both P1aNIa and P1a plasmids failed to enhance GFP accumulation (Figure III.7).

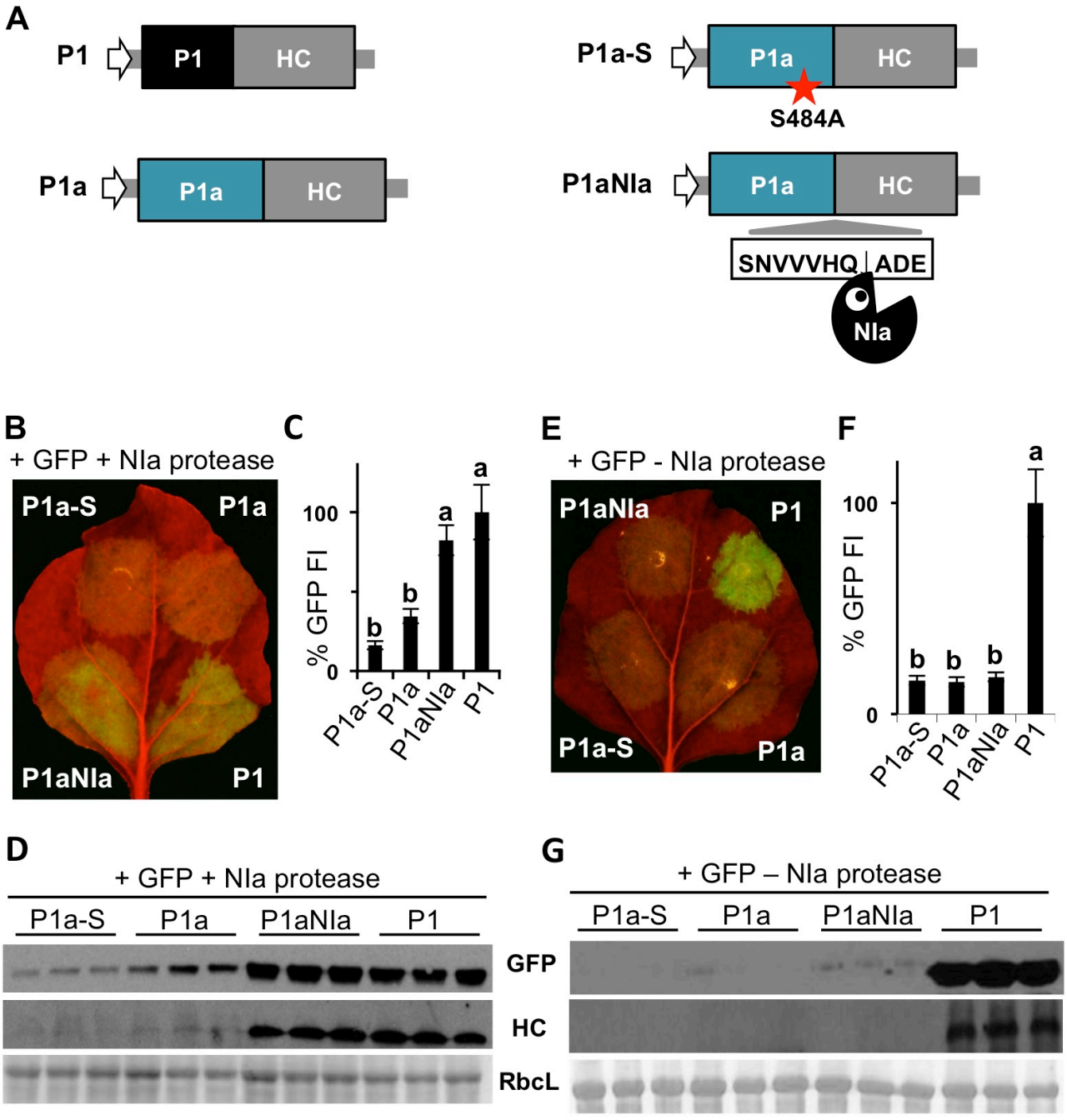

Figure III.7. HC silencing suppressor activity in different agroinfiltration constructs. (**A**) Schematic representation of the plasmids employed in transient agroinfiltration experiments. The stars mark serine to alanine mutations in the protease catalytic domain. (**B, E**) GFP fluorescence in a single leaf agroinfiltrated with four different constructs with (**B**) or without (**E**) NIapro. Pictures were taken on a blue light transilluminator at 6 dpa. (**C, F**) GFP FI of the agro-infiltrated leaf patches at 6 dpa quantified in a 96-well plate reader. Relative FI was plotted using P1 mean value equal to 100. Bar graph shows mean ± SD (n = 4 biological replicates); the difference between the results marked with different letters is statistically significant, $p < 0.01$, one-way Anova and Tukey's HSD test. (**D, G**) Anti-GFP and anti-HC western blot analyses of agroinfiltrated tissue at 6 dpa. Each lane corresponds to a sample pool from one/two agroinfiltrated plants; Ponceau red-stained blots showing RbcL were included as loading controls.

III.2.2. CVYV P1a and PPV P1 in *Cucumis sativus*

III.2.2.1. Viral accumulation of chimeric PPV bearing CVYV P1a is highly enhanced in local infections in *C. sativus*

Results in previous section indicate that incomplete CVYV P1a self-cleavage restricts the supply of PPV HC RSS activity and limits viral loads of PPV-derived clones carrying the CVYV P1a sequence in *N. benthamiana*. As a Type A P1 protein, CVYV P1a requires a still unidentified host factor to develop its proteolytiv activity (Rodamilans et al., 2013). Thus, the observed effects might be explained by the lack of compatibility between CVYV P1a and *N. benthamiana*, a non-permisive host for this virus. If this is the case, host factor-related defects caused by CVYV P1a should be amended when infection occurs in a natural host of CVYV, such as *C. sativus*.

PPV-derived chimera that were used to agroinoculate *N. benthamiana* plants were tested in cucumber (Figure III.8A). At 8 dpa, agro-inoculated leaves were analyzed by GFP fluorescence monitoring and anti-PPV CP immunoblotting to assess infectivity of the different constructs (Figure III.8B). Opposite to the results obtained in *N. benthamiana*, no appreciable GFP fluorescence or CP signal could be detected in plants infiltrated with the wild-type PPV construct, since it behaved as the negative control P1a protease mutant clone (PPV-P1aS). Leaves treated with chimeric viruses carrying CVYV P1a coding sequence (PPV-P1a and PPV-P1aNIa) showed GFP fluorescence and presented relevant amounts of CP, independent of the inclusion of an NIa cleavage site between P1a and HC (Figure III.8B). In agreement with Carbonell *et al*. (2012), leaves agroinfiltrated with PPV-P1aP1b construct appeared to accumulate a high CP amount (Figure III.8). In our experimental conditions, cucumber infections could only be detected locally suggesting that viral movement determinants are present somewhere else in potyviral genomes (Agbeci *et al.*, 2013;Vijayapalani *et al.*, 2012;Wei *et al.*, 2010).

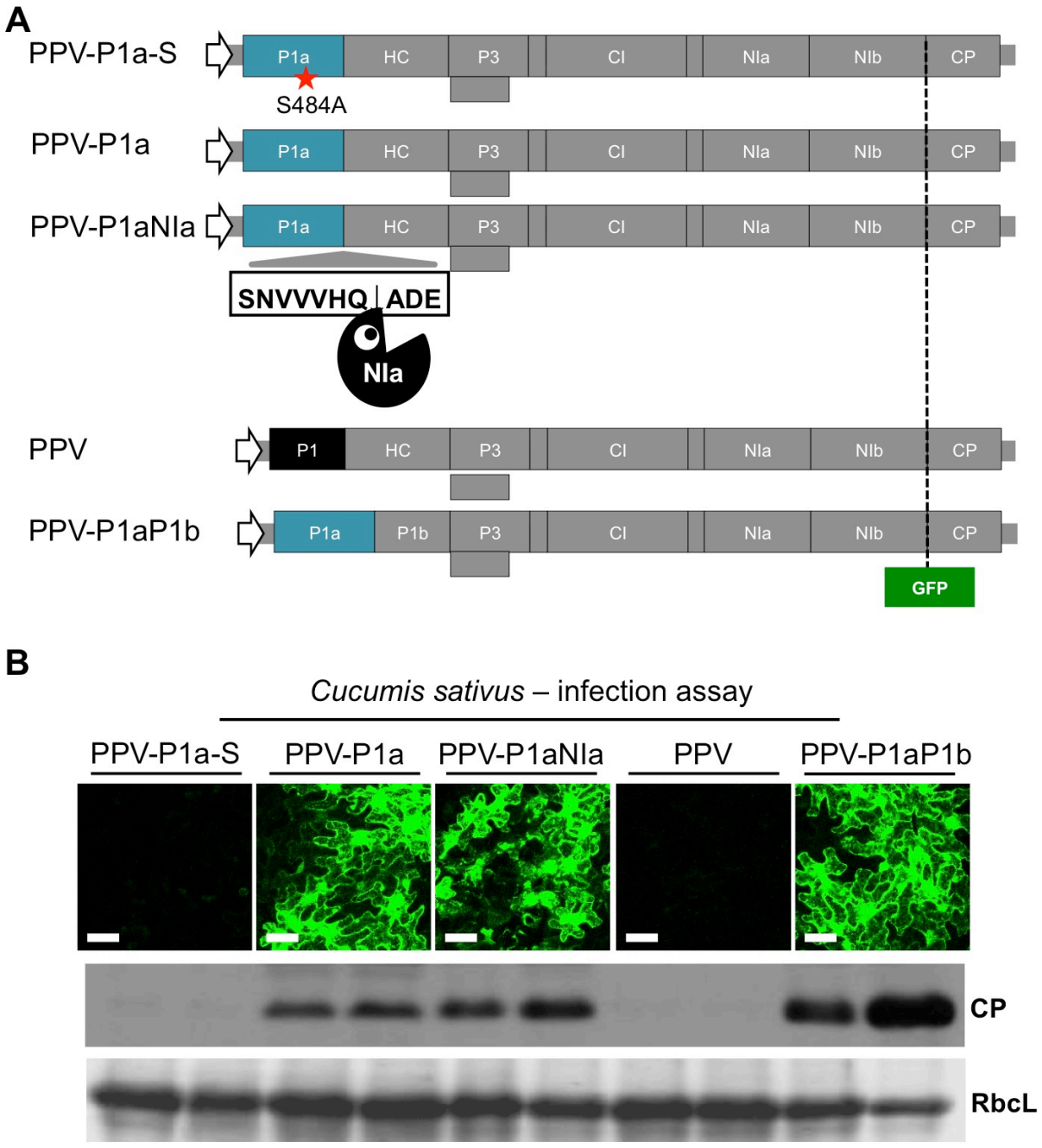

Figure III.8. Viral infection assays in *C. sativus* plants. (**A**) Schematic representation of the agro-inoculated chimeric cDNA clones. (**B**) PPV-derived chimeras were delivered to cucumber plants by agro-inoculation. At 8 dpa, confocal microscopy pictures were taken and local inoculated tissue was collected; scale bars, 100 µm. Viral accumulation was determined by anti-CP western blot. In all the panels, each lane corresponds to a single agroinfiltrated plant. The Ponceau red-stained blot showing the RbcL was included as loading control.

III.2.2.2. *cis*-supply of CVYV P1a effectively sustains PPV HC RSS activity in *C. sativus*

As shown in the previous section, in *N. benthamiana* plants the incomplete P1a processing impaired HC silencing suppressor activity (Figure III.7). To verify the host factor-dependency of

this result, previously designed construct P1a, which transiently expresses P1a-HC (Figure III.9A) was used to co-agroinfiltrate cucumber leaves with GFP. p19 and Φ were used as positive and negative controls, respectively. Western blot analysis of agroinfiltrated tissue at 6 dpa showed that co-expression of P1a-HC (sample P1a) sustained GFP accumulation to levels significantly above the empty vector control ($p < 0.05$), and similar to the p19 RSS positive control (Figure III.9B). The experiment was performed in parallel in *N. benthamiana* plants using the same agroinfiltration cultures (Figure III.9C). As in *C. sativus*, expression of the p19 plasmid was enough to sustain high levels of GFP in the agroinfiltrated leaves after 6 dpa. However, consistent with previous results (Figure III.7), no significant differences ($p > 0.05$) in GFP accumulation were detected between P1a construct and the empty vector (Figure III.9).

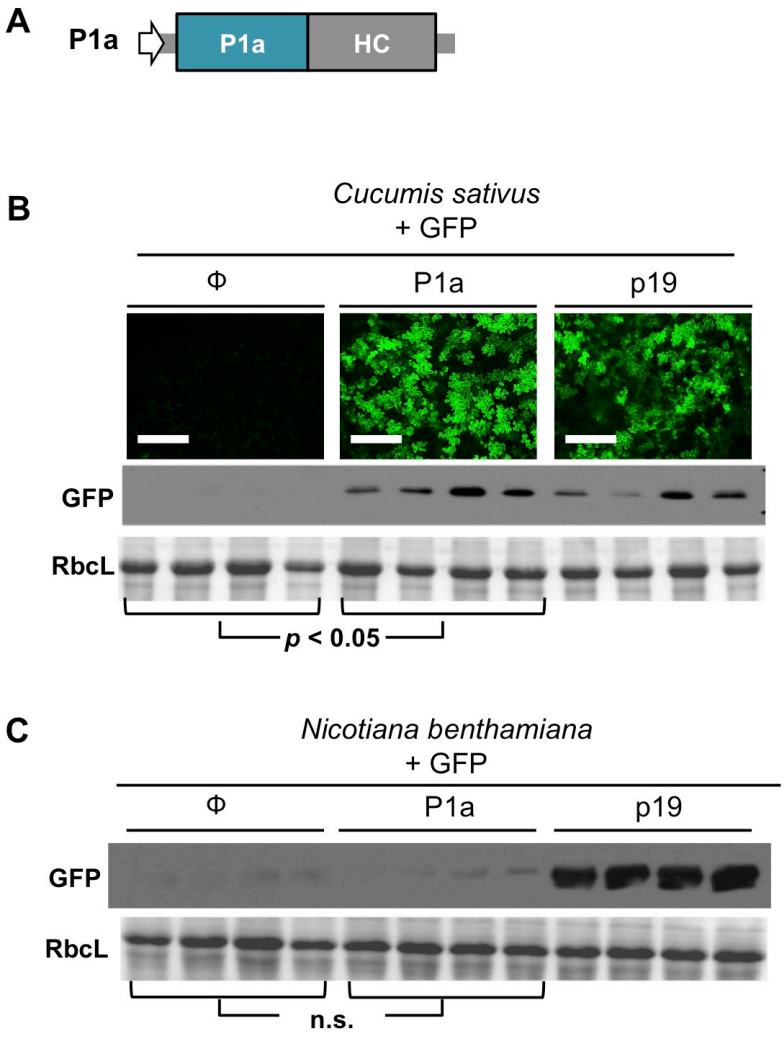

Figure III.9. Effect of P1a and HC on transient expression of GFP in *C. sativus* plants. (**A**) Schematic representation of the transiently expressed P1a plasmids. **(B and C)** Transient expression assays were done in cucumber (B) and, in a parallel experiment, in *N. benthamiana* plants (C). A GFP-expressing strain was co-

infiltrated with cultures containing an empty plasmid (Φ) or the P1a-HC-expressing construct P1a; a p19-expressing construct was used as positive control. Samples were analyzed at 6 dpa. Pictures of *C. sativus* samples were taken under a confocal microscopy; scale bars, 100 μm. (B). GFP accumulation was assessed by anti-GFP western blot analysis (B and C); the difference in GFP signal intensity between the indicated samples is statistically significant ($p < 0.05$) or not significant (n.s.), by Student *t*-test. In all the panels, each lane corresponds to a single agroinfiltrated plant. Ponceau red-stained blots showing RbcL were included as loading controls.

III.2.3. N-terminal deletion of P1 facilitates PPV replication in a non-permissive host

III.2.3.1 Delimitation of the antagonistic domain of PPV P1

Previous studies mapped the border of the C-terminal protease domain of PPV P1 between amino acids 165 and 169 (Pasin et al., 2014b). In contrast with the full-length P1, truncated proteases with the N-end at amino acid V164 (P1Pro) or R165 were active not only in WGE but also in RRL (Pasin et al., 2014b). To define the limits of the N-terminal sequences that interfere with the protease activity of PPV P1 in the absence of plant co-factors, a new set of N-terminal deletion P1 mutants were prepared fused to the N-terminal part of PPV HC (HC*). P1Pro and full-length P1 with mutations in their catalytic serines were included as negative controls (Figure III.10A). As already reported, V164 showed high processing capacity in both WGE and RRL with more than 50% of P1 detached from the HC part while full-length P1 presented high protease activity only when translated in WGE (>50%). Translation in RRL rendered P1 protein practically inactive in terms of processing capacity (<5%) (Figure 10B, C). As anticipated, the different P1 constructs presented very similar self-cleavage abilities in WGE, showing high processing ratios in all cases (>50%). On the other hand, when deletion mutants were translated into RRL, activity was gradually reduced until it reached the residual processing of full-length P1 (Figure III.10B, C). These findings show that there is not a well-defined antagonistic module at the N-terminal region of P1, but sequences of this region progressively contribute to prevent cleavage activity in the absence of plant co-factors.

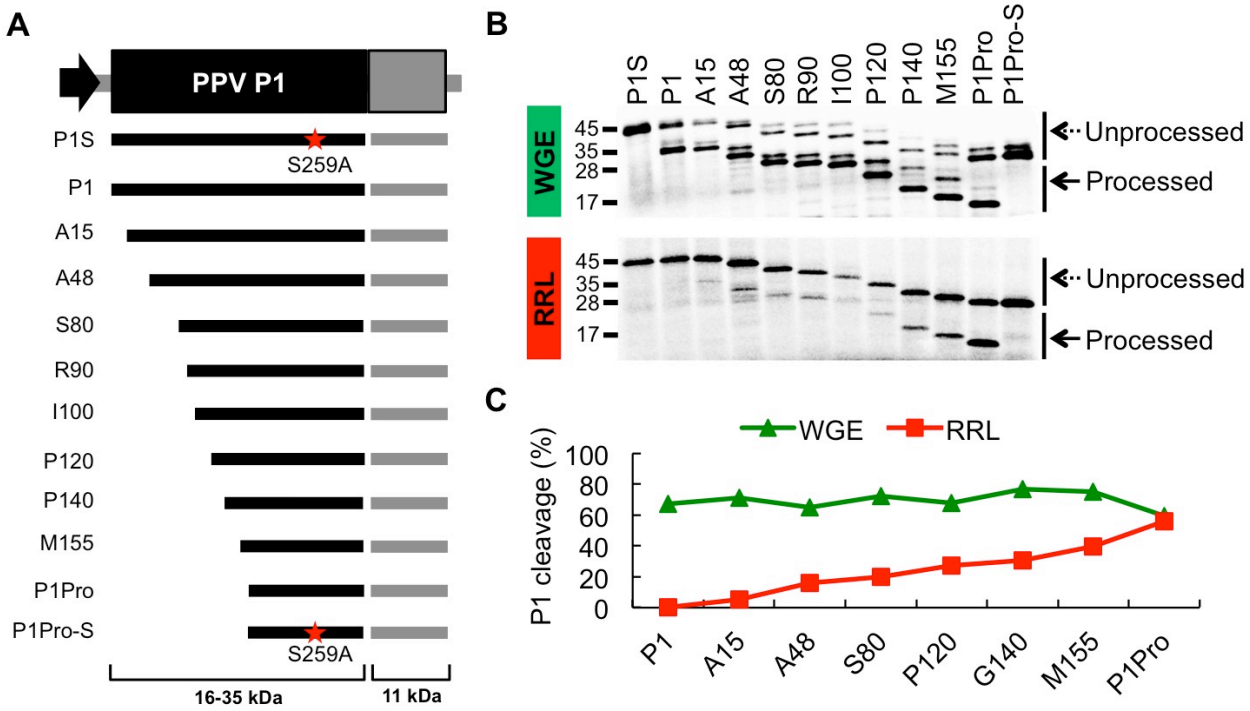

Figure III.10. Proteolytic activities of PPV P1 N-terminal deletion constructs in WGE and RRL *in vitro* translation systems. (**A**) Diagram shows DNA constructs with progressive deletions of PPV P1 coding sequences, amplified by PCR and subjected to *in vitro* transcription, and translation. Expected translation products and their molecular weights are displayed below. Truncation products are named according to the first amino acid (in addition to the initial methionine) that is maintained. Mutations in the catalytic serines of PPV P1 and PPV P1Pro (S259A) are marked with a red star. (**B**) Self-cleavage activity was evaluated in RRL and WGE *in vitro* translation mixtures. ^{35}S-labeled products were resolved by SDS-PAGE and detected by autoradiography. Processed and unprocessed products are marked (right); left, molecular weight markers. (**C**) Quantification analysis of the protease activity of the truncated P1 proteins. Plot was built based on P1 cleavage efficiency estimated by the program Quantity One. One hundred was assigned to maximum translation value, which includes processed and unprocessed products of P1 construct.

III.2.3.2 CVYV P1a and TuMV P1 proteolytic domains remain host factor dependent *in vitro*

Dependence on a plant factor(s) to develop their proteolytic activities is a distinctive feature of Type A P1 proteins, in contrast to Type B. PPV P1 protease domain detached from antagonistic N-terminal sequences can work independently of plant co-factors. Thus, it would be reasonable to anticipate that this property of PPV P1 protease domain could be shared by similar protease domains of other P1a-like proteins. To test this hypothesis, based on an alignment performed between PPV P1, CVYV P1a and TuMV P1, and using as reference the VELI motif (Valli et al., 2007), a series of deletion mutants were prepared intended to define plant co-factor-independent CVYV P1aPro and TuMV P1Pro (Figure III.11A). *In vitro* results showed that, contrary to our expectations, neither CVYV P1a nor TuMV P1 presented a protease domain active in both WGE

and RRL (Figure III.12). For both proteins, some of the tested deletions impaired protease activity in both WGE and RRL (R404 and L406 of CVYV P1a, Q223 and R230 of TuMV P1). Smaller deletions (K396, R398, E400 and S402 for CVYV, C218 for TuMV) allowed processing in WGE, but prevented cleavage in RRL indicating that these protease domains are functional, but rely on a plant factor(s) for their activity. This also suggests that the plant factor(s) interacts with the C-terminal part of P1a-like proteins of CVYV and TuMV.

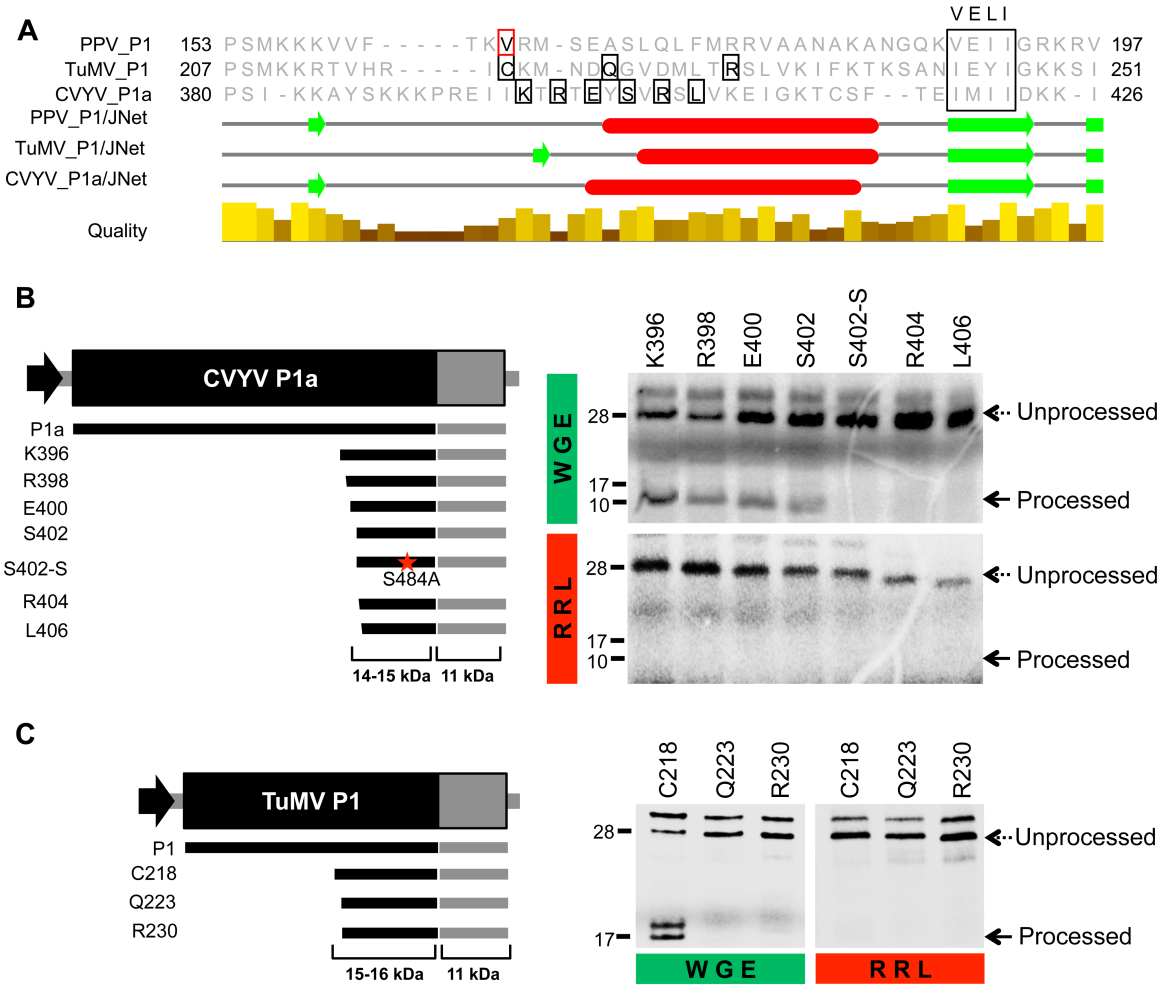

Figure III.11. Proteolytic activity of CVYV P1a and TuMV P1 minimal protease domains in WGE and RRL *in vitro* translation systems. (**A**) Partial alignment of PPV P1, TuMV P1 and CVYV P1a. The N-terminal amino acids of PPV P1Pro (V164) and of each construct tested for proteolytic activity are boxed in red and black, respectively. VELI motif is marked. Secondary structures were predicted using the JNet algorithm included in the Jalview package (Waterhouse *et al.*, 2009): α-helices, red ovals; β-sheets, green arrows; below, alignment conservation bars. (**B, C**) Diagrams show DNA constructs with progressive deletions of CVYV P1a (panel B) and TuMV P1 (panel C) coding sequences, amplified by PCR and subjected to *in vitro* transcription and translation. Expected translation products and their molecular weights are displayed below. Truncation products are named according to the first amino acid (in addition to the initial methionine) that is maintained. Mutation in the serine of CVYV P1a (S484A) is marked with a red star. Self-cleavage activity was evaluated in WGE and RRL *in vitro* translation systems. ^{35}S-labeled translation products were resolved by SDS-PAGE and detected by autoradiography.

Processed and unprocessed products are marked (right); left, molecular weight markers. Unaccounted extra bands might be originated from additional initiation events of translation.

III.2.3.3 A proper N-terminus is essential for plant host factor-independent activity of PPV P1pro

To verify if the singular processing capacities of PPV P1Pro among P1a-like proteins depend on specific features of its N-terminus, the first two amino acids of PPV P1Pro were mutated to alanines (PPV P1ProAA) and the processing capacities of the new mutant were compared with those of non-mutated PPV P1Pro (Figure III.12A). Mutants on the catalytic serines of P1Pro and P1ProAA (P1ProS and P1ProAAS, respectively) were included as negative controls. As it was shown before, P1Pro retained its proteolytic activity in WGE as well as RRL, however, PPV P1ProAA cleavage capacity was reduced more than 70% when assayed in RRL, while its activity in WGE remained almost unaltered (Figure 12B, C), indicating that when the N-terminus of PPV P1Pro is disturbed, it adopts a plant co-factor-dependent configuration similar to those of other protease domains of P1a-like potyvirid P1 proteins.

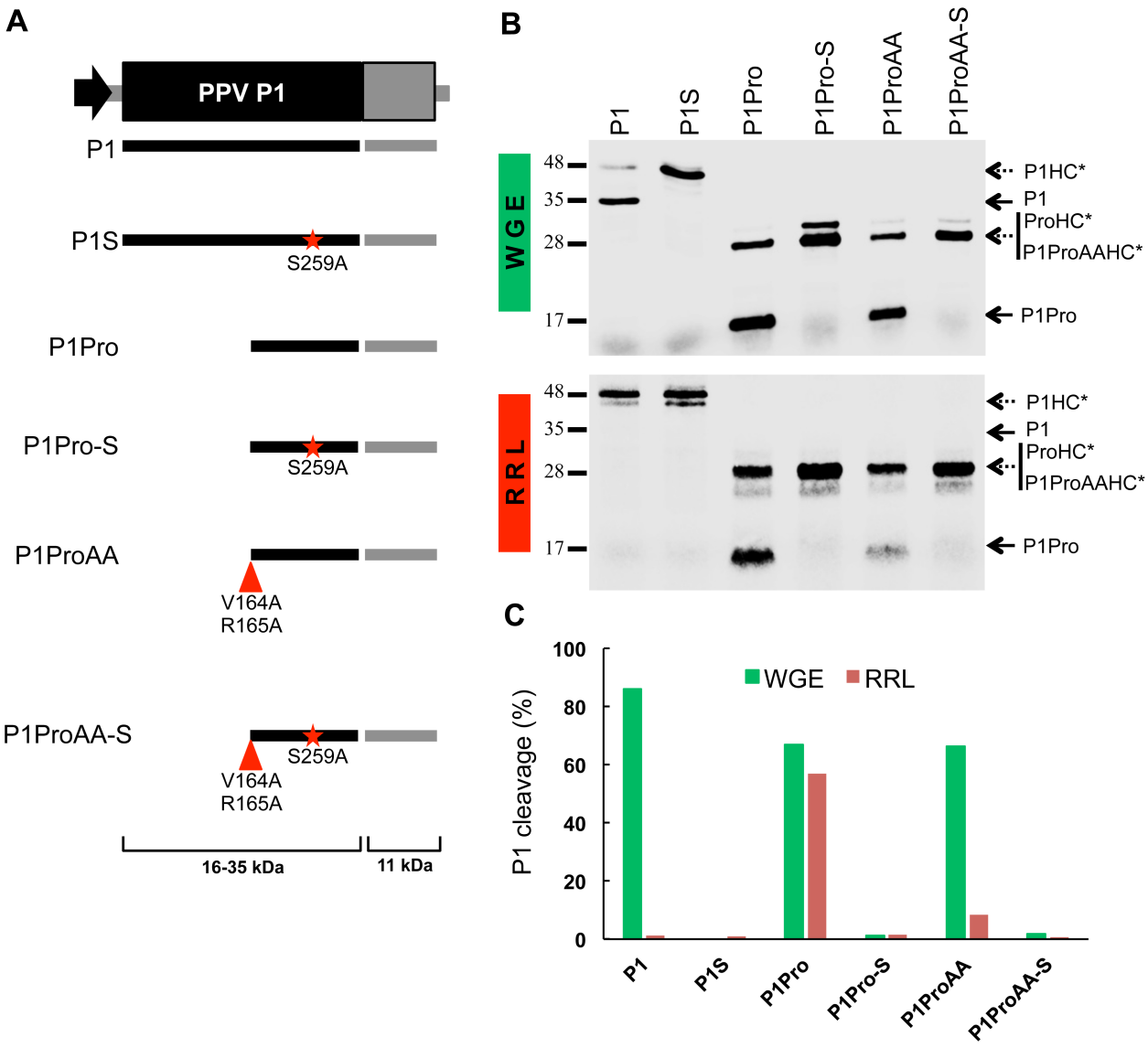

Figure III.12. Proteolytic activity of PPV P1Pro and P1ProAA truncated proteins in RRL and WGE *in vitro* translation systems. (**A**) Diagrams show DNA constructs of coding sequences of different PPV P1 variants amplified by PCR and subjected to *in vitro* transcription and translation. Expected translation products and their molecular weights are shown below. Mutations in the serine of the P1 active center (S259A) and in two residues at the N-terminus of the P1Pro domain (V164A and R165A) are indicated with a red asterisk and a red triangle, respectively. (**B**) Self-cleavage activity was evaluated in WGE and RRL *in vitro* translation mixtures. ^{35}S-labeled translation products were resolved by SDS-PAGE and detected by autoradiography. Processed and unprocessed polyprotein products are marked with solid and dotted arrows, respectively, (right); left, molecular weight markers. (**C**) Quantification analysis of the proteolytic activity of P1 variants, estimated by the program Quantity One. One hundred was assigned to maximum translation value, which includes processed and unprocessed products of P1 construct.

III.2.3.4 PPV P1Pro activity is not restricted by host specificity *in planta* and highly facilitates local PPV replication in *C. sativus*

PPV P1Pro presented plant co-factor-independent cleavage activity in *in vitro* assays and a chimeric PPV clone in which the P1 cistron was replaced by the coding sequence of P1Pro

showed faster replication rates at early times of infection and more severe symptoms in *N. benthamiana* plants (Pasin et al., 2014b). These results were attributed to an unrestricted ability by P1Pro to release itself from the polyprotein although this hypothesis was not tested. Experiments of local infection of PPV and CVYV P1a-containing PPV chimeras in cucumber described above suggested that in this host P1 could not detach itself effectively from HC rendering this protein inactive as RSS (Figure III.8).

To assess the host-independent performance of P1Pro in plants, an RSS assay by agroinfiltration was conducted in *C. sativus* and *N. benthamiana*, similar to the one performed with CVYV P1aPro. GFP was used as reporter of HC RSS activity in plants expressing P1ProHC, P1HC, P1SHC (mutant of P1HC on the catalytic serine of P1) and a control Φ (Figure III.13A). Fluorescence was monitored and GFP and HC accumulation was assessed by immunoblot analysis at 6 dpa. As anticipated, noticeable accumulation of HC and GFP was detected when P1HC was expressed in *N. benthamiana*, but not in *C. sativus* plants. In contrast, P1Pro released functional HC allowing high levels of GFP accumulation both in *N. benthamiana* and in *C. sativus* (Figure III.13B, C).

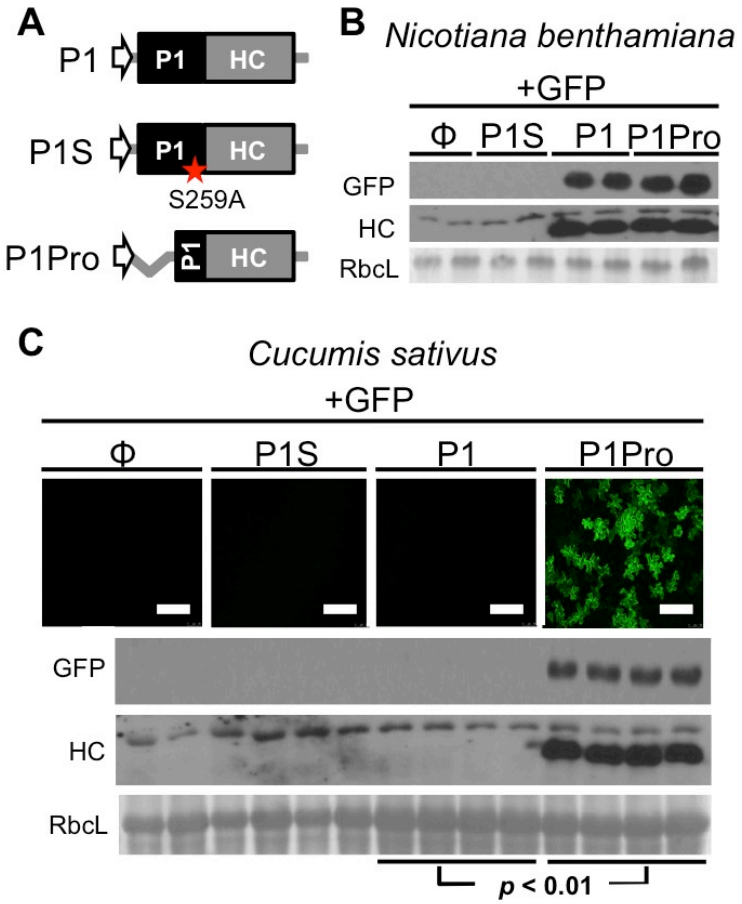

Figure III.13. Effects of upstream P1 sequences on PPV HC silencing suppressor activity in *N. benthamiana* and *C. sativus*. (A) Diagram of the agroinfiltrated constructs. Mutation in the catalytic serine of PPV P1 (S259A) is marked with a red star. **(B, C)** Transient expression by agroinfiltration was performed in *N. benthamiana* (B) and cucumber (C) leaves. A GFP-expressing construct was co-infiltrated with Φ or with P1-, P1S- and P1Pro-HC constructs. Samples were analyzed at 6 dpa. Images of *C. sativus* samples were taken under a confocal microscope; scale bars, 100 µm. GFP and HC accumulation were assessed by anti-GFP and anti-HC immunoblot assay. The Ponceau red-stained blot showing RbcL was included as loading control. The difference in GFP signal intensity between the indicated samples is statistically significant ($p < 0.01$) by Student's t-test ($n = 4$).

C. sativus is not a natural host of PPV and one of the factors limiting its compatibility might be the absence of a proper plant factor for the release of P1 from the rest of the polyprotein. To address whether the plant host factor-independent activity of P1Pro may contribute to overcome host-specific restrictions to viral infection, chimeric virus PPV-P1Pro was used to inoculate *C. sativus* and compare with PPV-inoculated plants (Figure III.14A). *N. benthamiana* plants were used as positive control of unrestricted infection. To verify that the lack of infectivity of PPV in *C. sativus* was driven by a defective HC RSS activity due to limited P1 self-cleavage, p19 was included in some of the infiltrations to complement RSS defects.

Confocal microscopy analysis of cucumber inoculated leaves at 7 dpa showed presence of GFP in all viral samples co-infiltrated with p19. When p19 was absent from infiltration mixtures, GFP could only be observed in PPV-P1Pro samples (Figure III.14B). This result suggests that, in cucumber, free HC able to support PPV infection is liberated by P1Pro but not by full-length P1. This assumption was confirmed by immunoblot analysis using anti-CP and anti-HC antibodies. CP was observed in all samples co-agroinfiltrated with p19, but released HC could only be detected in PPV-P1Pro samples, independently of the presence or absence of p19 (Figure III.14C). No systemic infection could be detected in any of the inoculated cucumber at 21 dpa. In *N. benthamiana* plants, where both P1 and P1Pro are equally functional, CP and HC could be detected in all viral clone samples (Figure III.14C).

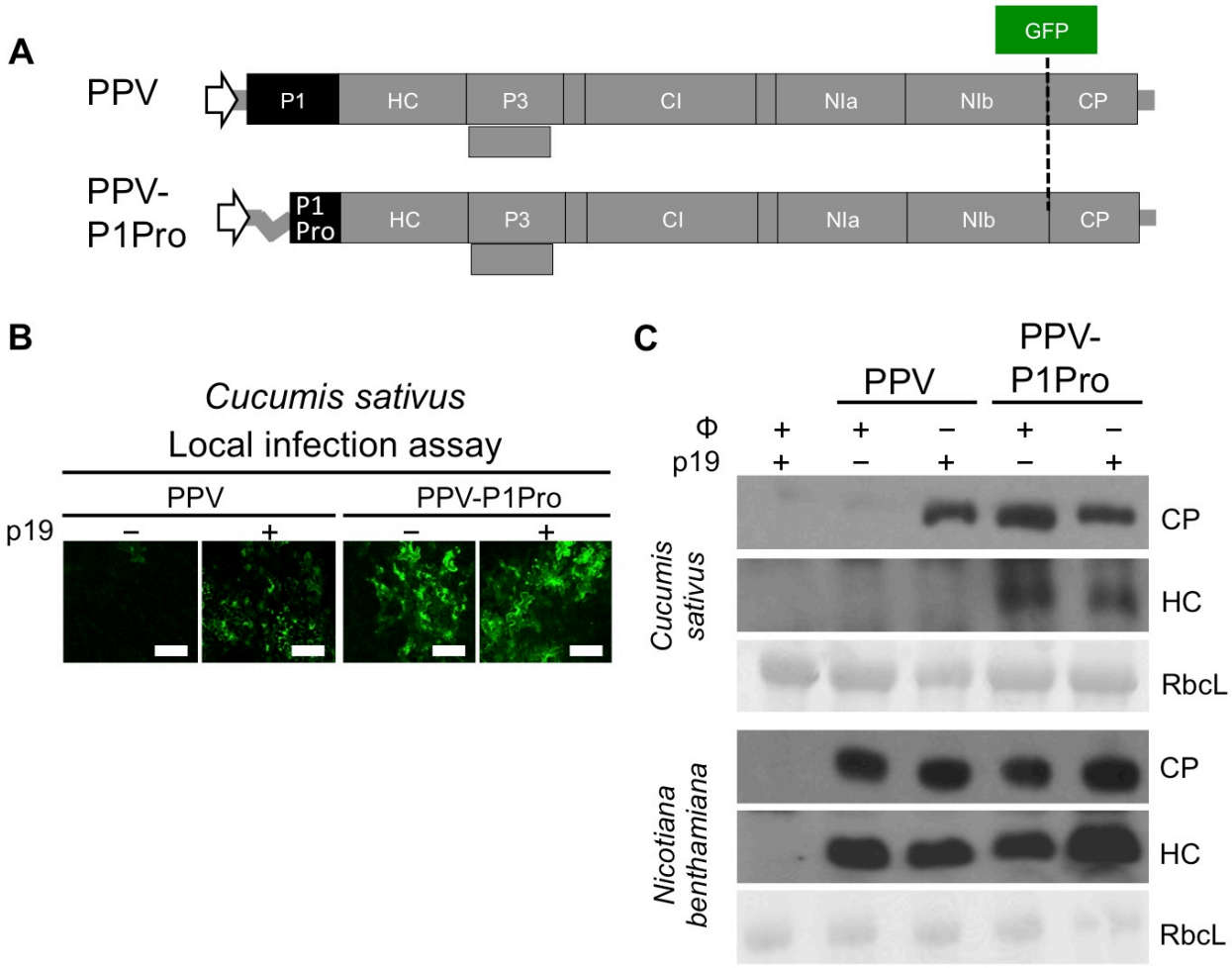

Figure III.14. Effects of N-terminal truncation of P1 on PPV infection in cucumber leaves. (**A**) Diagrammatic representation of PPV and PPV-P1Pro cDNA clones. These clones were delivered to *C. sativus* and *N. benthamiana* plants by co-agroinfiltration with Φ or p19 silencing suppressor; Φ and p19 constructs were co-agroinfiltrated together as a negative control. (**B**) At 7 dpa, confocal microscopy pictures of cucumber inoculated leaves were taken; scale bars, 100 µm. (**C**) Viral CP accumulation and release of HC was assessed by immunoblot assays in cucumber (top panel) and *N. benthamiana* (bottom) inoculated leaves. Each lane corresponds to a pool of four agroinfiltrated plants. Ponceau red-stained blots showing RbcL as loading controls.

The RNA-dependent RNA polymerase activity of the potyviral NIb protein is essential for viral RNA replication, and this activity is abolished by removal of the GDD catalytic motif (Guo *et al.*, 1999). To further demonstrate PPV-P1Pro replication in cucumber inoculated leaves, an independent experiment was performed including a PPV-P1Pro replication-defective mutant lacking the GDD motif (PPV-P1Pro$_{\Delta GDD}$) (Figure III.15A). Immunoblot analysis using anti-CP antibody shows CP accumulation only in cucumber plants agroinfiltrated with PPV-P1Pro (Figure III.15B). RT-qPCR analysis of positive and negative viral genome strands confirmed PPV-P1Pro replication in these leaves (Figure III.15C).

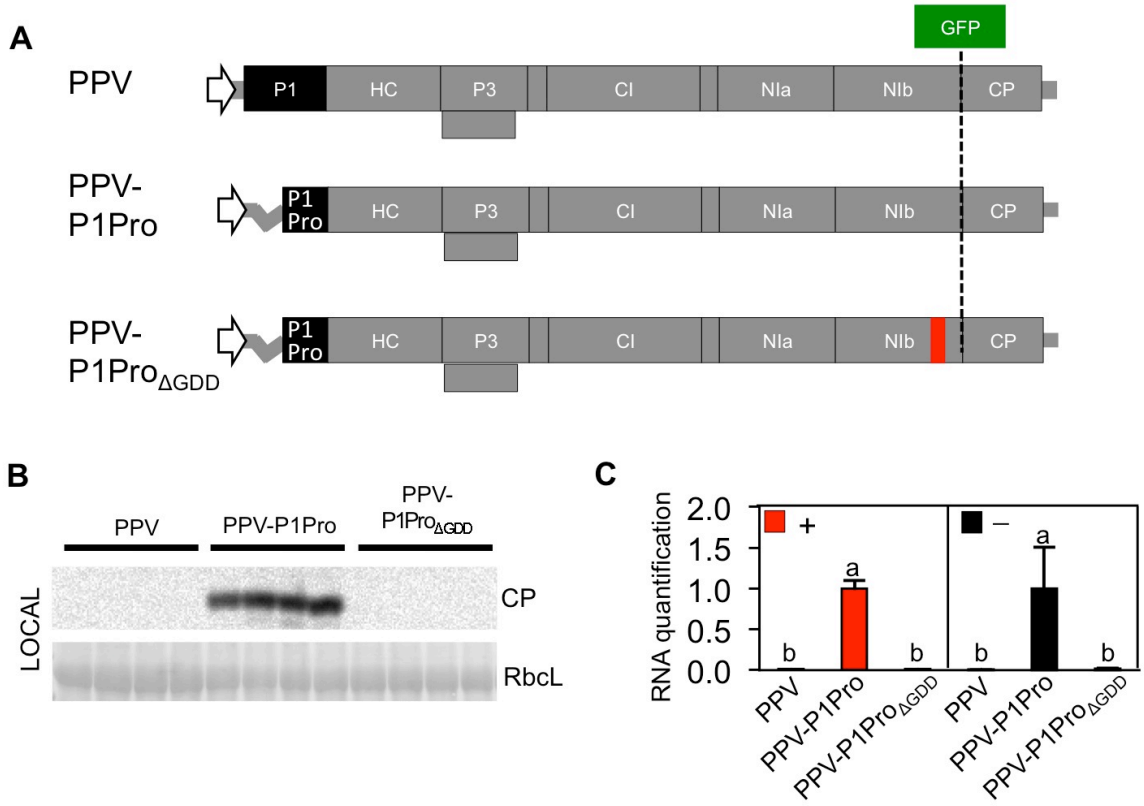

Figure III.15. Analysis of viral RNA replication in the inoculated leaves of *C. sativus*. (**A**) Diagram of the agroinfiltrated constructs. In the replication-defective mutant, NIb deleted amino acids that render replicase inactive are marked with a red line. (**B**) At 7 dpa, local viral accumulation was assessed by anti-CP immunoblot assay. Each lane corresponds to one plant sample. The Ponceau red-stained blot showing RbcL was included as loading control. (**C**) RT-qPCR analysis of accumulation of viral RNA of positive (left) and negative (right) polarity. Data were quantified relative to the average value of PPV-P1Pro. Bars show mean ±SD (n=4); letters indicate $p < 0.01$, one-way Anova and Tukey's HSD test.

III.2.3.5 Protease domain of CVYV P1a (P1aPro) is host-dependent *in planta* behaving as full-length CVYV P1a

In vitro tests performed with CVYV P1a allowed the identification of a minimal protease domain (P1aPro) that remains active in WGE, similar to the one defined for PPV P1. This domain, however, cannot perform autocatalytic processing in RRL suggesting a host factor(s) dependency, not observed in the case of the P1 protease domain (Figure III.12). To answer the question whether these *in vitro* results can be extrapolated to an *in planta* system, transient co-agroinfiltration experiments were performed in *N. benthamiana* and *C. sativus* hosts using expression constructs previously tested (P1 and P1a) as controls (Figure III.16A), and using newly designed constructs P1aPro (expressing P1aProHC) and P1aProS (the S484A proteolytically inactive mutant version of P1aPro) (Figure III.16A). All constructs were co-agroinfiltrated with a GFP expressing plasmid.

At 6 dpa, in *N. benthamiana* plants GFP fluorescence could only be detected in plants agroinfiltrated with P1 construct (Figure III.16B). Immunoblot analysis with anti-GFP and anti-HC confirms this. As in previous experiments performed in *N. benthamiana* (Figure III.7) leaves agroinfiltrated in this new experiment with P1a construct did not show any GFP fluorescence and no HC or GFP expression could be observed by western blot analysis. Constructs bearing the protease domain of CVYV P1a (P1aPro and P1aProS) presented similar negative result (Figure III.16B). A change of host to *C. sativus* produced a very different outcome regarding GFP and HC expression. In this case, GFP fluorescence was analyzed by confocal microscopy and GFP and HC expression was assessed by immunoblot assay as before. Plants agroinfitrated with P1 showed no presence of GFP or HC. In contrast, consistent with results described above (Figure III.9), plants treated with P1a presented GFP fluorescence, and both GFP and HC expression, as estimated by western blot analysis (Figure III.16C). Plants agroinfiltrated with P1aPro presented a similar result. GFP fluorescence could be detected by confocal microscopy and GFP and HC accumulation was also observed. Protein yields were comparable to the ones obtained with P1a construct. P1aProS agroinfiltrated cucumber plants showed the same negative result as in the case of *N. benthamiana* plants (Figure III.16C).

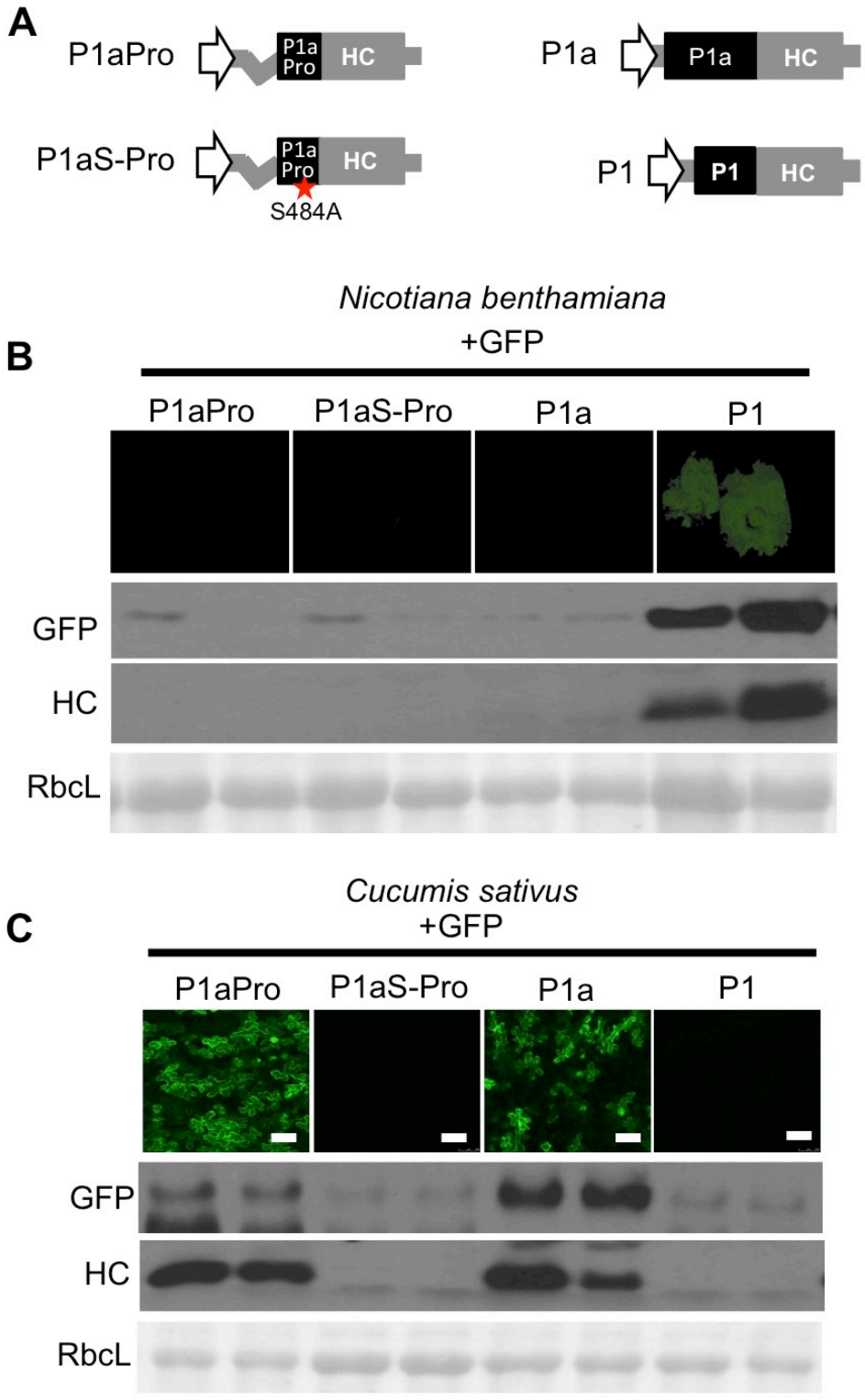

Figure III.16. P1aPro processing is host-dependent *in planta*. (A) Diagrammatic representation of agroinfiltrated clones. **(B, C)** Transient expression by agroinfiltration was performed in *N. benthamiana* (B) and *C. sativus* (C) plants. A GFP-expressing plasmid was co-infiltrated with the indicated constructs. Inoculated tissue was analyzed at 6 dpa. GFP photos of *N. benthamiana* were taken on a blue light transilluminator. Images of *C. sativus* samples were taken under a confocal microscope; scale bars, 100 µm. GFP and HC accumulation were assessed by anti-GFP and anti-HC immunoblot assays. Ponceau red-stained blots showing RbcL were included as loading control.

III.2.3.6 Deletion of the N-terminal antagonistic domain of P1 does not boost systemic infection in *Prunus* hosts

Out of the ten recognized PPV strains, D, M and Rec are the three major ones, sharing ~95% genome identity and displaying variable pathogenicity features regarding adaptation to specific hosts (García et al., 2014;Sihelská *et al.*, 2017). The isolate PPV-D, which belongs to the D strain is infectious in *Prunus* seedlings but it is only able to infect locally in *Nicotiana* species (Salvador et al., 2008a;Carbonell et al., 2012). On the other hand, PPV-R, another D-type isolate has been adapted to replicate and move efficiently causing systemic infections in several *Nicotiana* species such as *N. benthamiana*, but it is only able to infect locally in *Prunus* species (Salvador et al., 2008a). A chimeric virus carrying the 5´ region of PPV-D and the 3´region of PPV-R (PPV-5´BD) can infect herbaceous and woody hosts locally and systemically (Maliogka et al., 2012;Salvador et al., 2008a). In *C. sativus*, deletion mutant PPV-P1Pro was able to overcome host incompatibilities and boost local infection compared to wild type PPV. Considering the observed differences between PPV-R infection in *Nicotiana* and *Prunus* hosts, it is likely that part of these incompatibilities could derive from a similar host factor restriction mechanism. To test this hypothesis, PPV and PPV-P1Pro were used to agroinoculate seedlings of *Prunus domestica*, *Prunus persica* and *Prunus avium*. PPV-5´BD was used as positive control of infection in these hosts (Figure III.17A). *N. benthamiana* plants were used as positive control of unrestricted infection (not shown).

At 10 dpa, plants were analyzed checking GFP fluorescence of inoculated and upper non-inoculated leaves. Samples were taken and immunoblot assays using anti-CP and anti-HC were performed. In *P. persica* seedlings, GFP fluorescence could be detected in the inoculated leaves of samples challenged with PPV and PPV 5´BD. GFP presence in the upper systemic tissue was only observed in PPV 5´BD inoculated trees (Figure III.17B). Images were confirmed by immunoblot results (Figure III.17C, D). Similar result was observed in *P. domestica* (Figure III.17E, F, G). *P. avium* is known not to be a suitable host for PPV isolates of the D strain (Figure III.17H, I, J); in agreement with that, none of the *P. avium* seedlings showed signs of infection in the upper non-inoculated leaves. Inoculation of trees with PPV-P1Pro did not facilitate systemic infection in any of the species tested. Moreover, local infection was significantly reduced in all hosts compared to PPV.

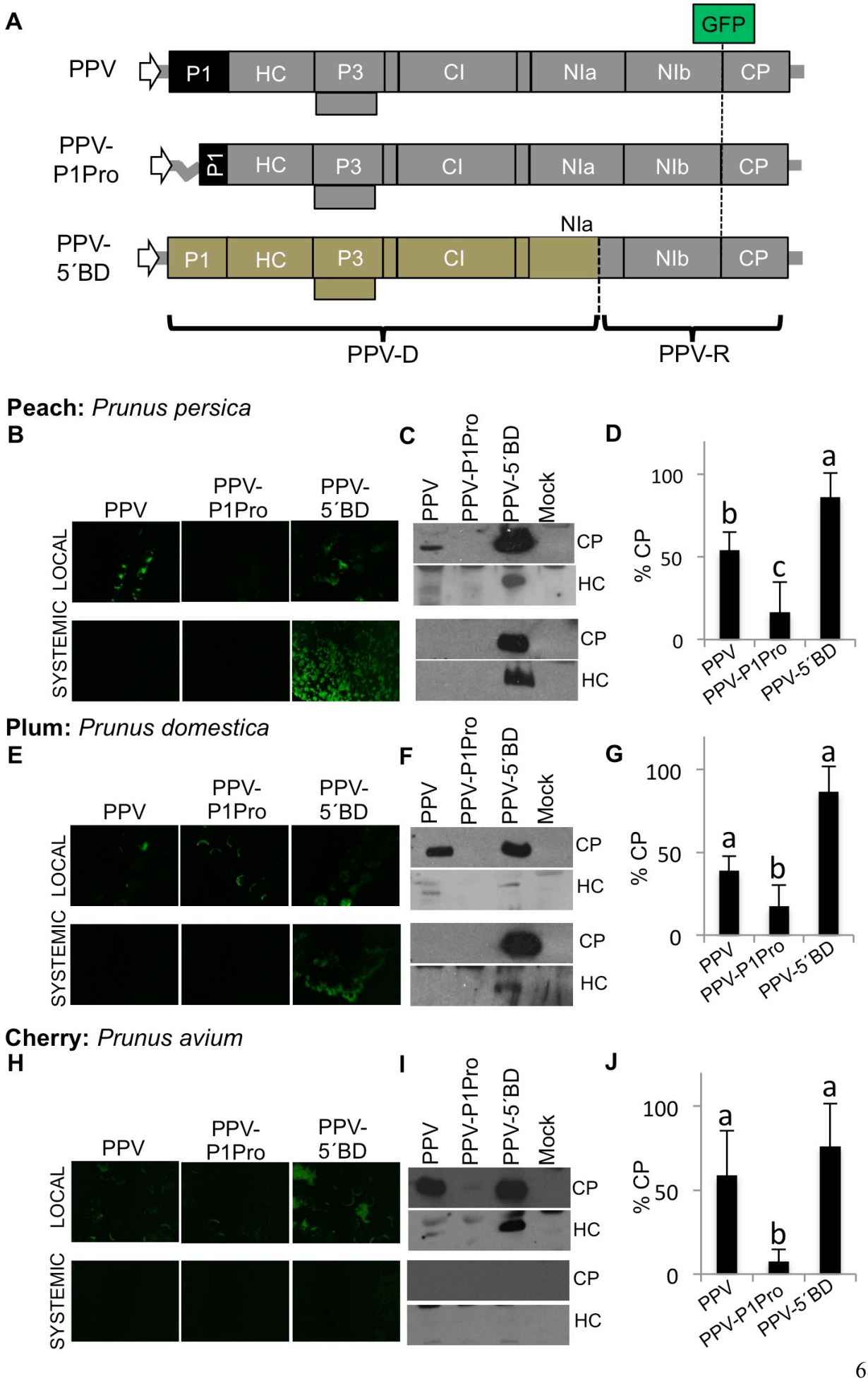
A
GFP
PPV
P1
HC
P3
CI
NIa
NIb
CP
PPV-P1Pro
PPV-5´BD
PPV-D
PPV-R
Peach: Prunus persica
B
C
D
PPV
PPV-P1Pro
PPV-5´BD
Mock
LOCAL
SYSTEMIC
CP
HC
% CP
Plum: Prunus domestica
E
F
G
Cherry: Prunus avium
H
I
J

Figure III.17. Effect of PPV P1 N-terminal deletion in the viral infection of *Prunus* seedlings. (A) Schematic representation of the three different viral constructs used. **(B, E and H)** PPV, PPV-P1Pro and PPV-5'BD were delivered to *Prunus* seedlings by agro-inoculation. Empty vector (mock) was used as a negative control. At 10 dpa, GFP pictures of the inoculated and upper leaves were taken from peach (B), plum (E) and cherry (H) leaves under an epifluorescence microscope. **(C, F and I)** Both local and systemic tissue from the inoculated peach (C), plum (F) and cherry (I) seedlings were collected to assess viral accumulation by anti-CP and anti-HC immunoblot assay. Each lane corresponds to a pool of four agroinfiltrated plants. **(D, G and J)** Quantification of the intensity of the CP signals from peach (D), plum (G) and cherry (J) inoculated leaves. The difference between the results marked with different letters is statistically significant according to one-way Anova and Tukey's HSD test. ($p<0.01$, $n=4$, using two independent *Agrobacterium* cultures).

III.3. Purification and identification of host co-factor(s) of PPV P1 protease

Characterized more than twenty years ago as a proteinase, P1 was nonetheless the last potyviral endopeptidase identified (Carrington et al., 1990;Verchot et al., 1992;Verchot & Carrington, 1995b). Being a non-essential protein, its release from the polyprotein is indispensable for virus infection (Verchot & Carrington, 1995a). Computational analysis of P1 potyviral proteins showed its great variability, both in length and in amino acid sequence, and its diversification in potyviral species was associated with host specialization (Valli et al., 2007). This host-dependent relationship was described early on, by *in vitro* translation assays (Verchot et al., 1992) and corroborated by other experimental systems in more recent reports (Rodamilans et al., 2013). In addition, previous works (Carbonell et al., 2012;Pasin et al., 2014b) and results presented in the precedent sections of this thesis help for a better understanding of P1 processing and its implication in host range determination. Taken together, all these data indicate that host factor(s) identification is key to understand potyviral infection process and support the notion that P1 leader proteases and the plant host factor(s) involved in its proteolytic activity are ideal targets for antiviral therapies.

Recent works studied possible plant protein interactors with TEV (Martínez & Daròs, 2014), TuMV (Pan, 2016) and Soybean mosaic virus (SMV) P1 proteins (Shi *et al.*, 2007). Data obtained were useful to proposed extra roles that P1 might be having during viral infection, such as its involvement in protein translation. None of these reports, however, aimed at specifically finding the factor(s) responsible for the autocatalytic cleavage of P1 proteins. Considering auto-cleavage as the most relevant function described for P1 leader proteinases, this last part of the thesis intends to identify the host factor(s) responsible for P1 self-processing using a biochemical approach that combines protoplast extract preparation from *N. tabacum* cells, classical purification techniques coupled to *in vitro* translation assays and state-of-the-art mass spectrometry analysis for candidate identification.

III.3.1. Preparation of an extract from *N. tabacum* BY-2 cells that can complement PPV P1 proteolytic activity in RRL translation system

III.3.1.1 PPV P1 deficient autocatalytic activity in RRL can be restored by addition of a small percentage of WGE to the translation mixture

As previously reported (Verchot et al., 1992), TEV P1 expressed in RRL was not able to perform autocleavage, but addition of small amounts of WGE to the reaction mixture prior to translation restored the catalytic ability of P1. This finding led the authors to conclude that some component in the WGE was complementing the RRL deficient mixture. WGE denaturation before addition into the RRL translation system abolished the observed complementation effect supporting the idea that the factor(s) responsible for this was a protein.

To test the ability of a small amount of WGE to complement PPV P1 processing after *in vitro* translation in RRL, an experiment similar to the one described for TEV P1 was performed (Figure III.18). The previously tested P1HC* mRNA was expressed in RRL and WGE extracts as controls. As anticipated, a band corresponding to the unprocessed P1HC* product was observed in RRL and WGE samples, however, a band matching the size of the processed PPV P1 product was observed only when translation was performed in WGE. Addition of 20% WGE to the RRL translation mixture allowed the detection of two bands corresponding to the processed and unprocessed products indicating P1 proteolytic complementation. Denaturation of WGE prior to RRL translation disturbed the positive effect observed on PPV P1 cleavage obtaining a similar result as when the transcript was expressed in RRL alone, suggesting the protein nature of the plant factor(s) (Figure III.18).

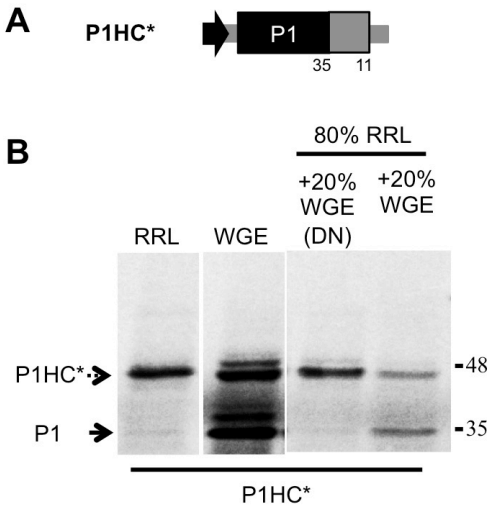

Figure III.18. Self-processing activity of PPV P1 in RRL supplemented with WGE. (**A**) Diagram of the DNA construct of the coding sequence of the leader protease of PPV, amplified by PCR and subjected to *in vitro*

transcription and further *in vitro* translation, T7 RNA polymerase promoter drives the PPV 5'UTR and the coding sequences of PPV P1 (black box) and 97 N-terminal amino acids of PPV HC (HC*, grey box) followed by a stop codon and PPV 3'UTR. Molecular weights of the expected translation products are displayed below (**B**) Proteolytic activity assay of PPV P1 in different translation mixtures. ^{35}S-labeled products were resolved by SDS-PAGE and detected by autoradiography. Processed and unprocessed products are marked (left); right, molecular weight markers.

III.3.1.2 A protein extract from *N. tabacum* BY-2 cells is able to complement PPV P1 processing allowing autocatalytic activity in RRL translation system

BYL extract preparation process was detailed in Materials and Methods, section II.11. Briefly, *N. tabacum* BY-2 cells were grown in MS medium for 4-5 days (Figure III.19A). Cells were collected by centrifugation and cell walls were removed by enzymatic digestion and resulting protoplasts were set up on top of a Percoll gradient for centrifugation and separation of evacuolated and non-evacuolated protoplasts (Figure III.19B). Since vacuoles might contain contaminants that could hamper following *in vitro* applications (Hara-Nishimura & Hatsugai, 2011;Murota et al., 2011) both kinds of protoplasts were processed separately from this point forward to obtain two kinds of BYL extracts, BYL-P and BYL-EP (from complete and evacuolated protoplasts, respectively). Protoplasts and evacuolated protoplasts from the Percoll gradient were resuspended in TR buffer (Murota et al., 2011) and disrupted with a Dounze homogenizer. To remove rests of Percoll and other contaminants, extracts were purified on PD-10 columns using TR buffer for elution. Protein complexity was evaluated by SDS-PAGE analysis (Figure III.19C). BYL-P and BYL-EP were stored at -80 ºC for further *in vitro* testing.

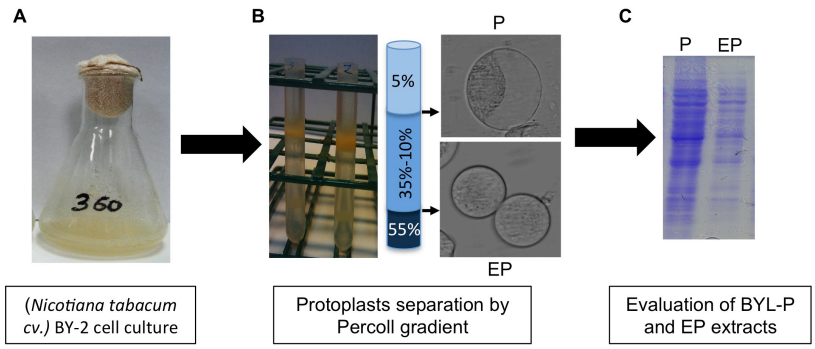

Figure III.19. Protein extract preparation from *N. tabacum* BY-2 cells. (**A**) Cell culture prior to protoplasts preparation. (**B**) Percoll gradient after centrifugation with the two layers containing complete (top of the gradient) and evacuolated protoplasts (bottom of the gradient). (**C**) SDS-PAGE evaluation of the protein complexity of BYL-P and -EP extracts (2 µl).

BYL-EP extract was used in the following experiments to test: i) its ability to complement P1 protease activity in RRL (Figure II.20B); ii) the protein nature of the factor(s) by denaturation of the extract prior to complementation (Figure III.20C) and iii) the effect in P1 cleavage of adding increasing amounts of BYL-EP (Figure III.20D, E) in the RRL reaction mixture. BYL-P extract was used to test the deleterious effect on *in vitro* translation of the presence of vacuoles during extract preparation (Figure III.20F).

For the initial test, P1HC* construct used previously (Figure III.20A), was employed for *in vitro* translation in RRL supplemented with either BYL-EP extract or TR buffer. P1SHC* mRNA carrying a mutation in the catalytic serine of PPV P1 was used as negative control. As can be observed in Figure III.20B, one band corresponding to the unprocessed product P1HC* could be detected, but no band matching the size of single P1 protein was observed. Same result was obtained in the case of the negative control in which only a band of the size corresponding to the unprocessed P1SHC* product is detected. On the other hand, when P1HC* was supplemented with BYL-EP extract, a clear product of the approximate size of P1 is observed together with a similar product of a size matching the unprocessed polypeptide. This result indicates that BYL-EP can complement PPV P1 processing in RRL. Following this, P1HC* transcript was used for translation in RRL supplemented with either BYL-EP extract or the same extract previously denatured by heating it 5 min at 95 ºC [BYL-EP(DN)]. The result shows that after denaturing the extract, the band previously observed corresponding to P1 processed product is no longer present suggesting, as was the case for WGE, that the factor(s) found in BYL-EP is of protein origin (Figure III.20C). Finally, P1HC* transcript was translated again in RRL complemented with increasing amounts of BYL-EP extract. Maximum percentage of BYL tested was 60% with respect to the total amount of RRL in the translation mixture. Samples with lower amounts of BYL were compensated with the proportional amounts of TR buffer. The result indicates a direct relationship between the amount of P1 processed band observed and the amount of extract added to RRL, further supporting the hypothesis of a factor(s) of protein origin and indicating the amount of BYL extract to be used in further complementation tests (Figure III.20D).

BYL-P extract was tested as P1 proteolytic activity complement in RRL translation system. P1HC* (Figure III.20A) was used as mRNA. As positive control same construct was translated in RRL complemented with BYL-EP, previously tested. Contrary to what was anticipated, BYL-P extract did not carry enough contaminants to restrict translation of P1HC* product, as a band corresponding to the unprocessed product can be observed. Besides, BYL-P presented enough P1 related-factor(s) to allow detection of a band corresponding to the cleaved P1 product (Figure III.20F).

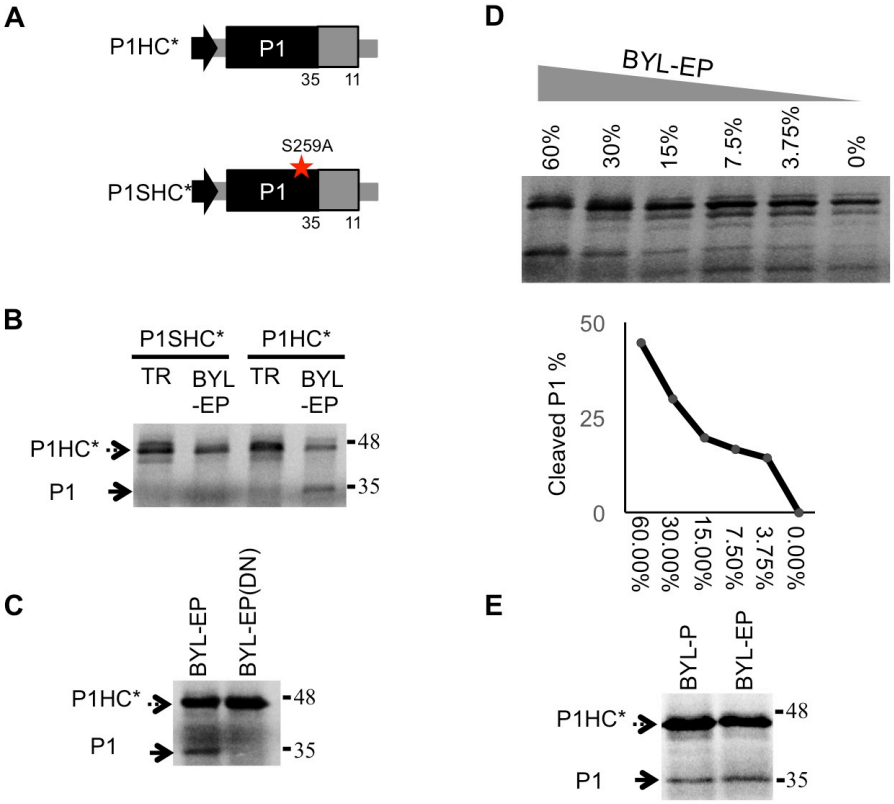

Figure III.20. PPV P1 catalytic activity assay in RRL supplemented with different BYL extracts. (**A**) Diagrams of DNA constructs of the coding sequences of leader proteases of PPV, amplified by PCR and subjected to *in vitro* transcription and further *in vitro* translation, T7 RNA polymerase promoter (black arrow) drives the PPV 5'UTR and the coding sequence of the P1 protease (black box) and 97 N-terminal amino acids from HC (HC*, grey box) followed by a stop codon and PPV 3'UTR. Mutation in the serine of the catalytic center (S259A) is marked with a red star. Molecular weights of the expected translation products are displayed below. (**B**) RRL *in vitro* translation with and without BYL-EP (20 % of the reaction mixture) supplement. TR buffer was used as negative control. (**C**) RRL *in vitro* translation supplemented with BYL-EP (20 %) before and after denaturation. (**D**) P1 cleavage *in vitro* assay (upper panel) and densitometric quantification analysis (lower panel) using increasing amounts of BYL-EP. Supplements were completed up to 60% with TR buffer. Plot was built based on P1 cleavage efficiency estimated by the program Quantity One. One hundred was assigned to maximum translation value, which includes processed and unprocessed products of P1 construct. (**E**) RRL *in vitro* translation supplemented with BYL-P or BYL-EP extracts (20%). In B, C and E, processed and unprocessed products (left) and molecular weight markers (right) are signaled.

III.3.2. Purification of the P1 co-factor(s) from BYL extract

The following purification experiments were conducted using a combination of both extracts BYL-P and BYL-EP (now termed BYL) to increase protein yield and raise factor(s) concentration in the final purified sample. Three separate trials using different BYL extracts were performed.

III.3.2.1 Anion exchange and size exclusion chromatography

Details on the purification process are described in Materials and Methods, sections II.12 and II.13.

For the initial purification trial, a commercial 1-ml HiTrap Q HP column (Amersham) coupled to an FPLC device was used for anion exchange chromatography. Column was equilibrated with TR buffer (200 mM KAc). Two independent BYL extracts, BYL-1 and BYL-2, were combined (9.5 mg of total protein) and salt concentration was adjusted to 200 mM KAc before loading the column. Step elution of the proteins was performed by washing the column with TR buffer in varying amounts of KAc, from 200 mM to 1 M. Flow-through and eluted fractions were collected, concentrated (Q-cFT and Q-cEF, respectively) and dialyzed. Protein complexity of the samples was analyzed by SDS-PAGE (Figure III.21A). Fractions were then tested for complementation of P1 cleavage in RRL translation system using P1HC* mRNA previously described. P1SHC* catalytic mutant was included as negative control (Figure III.21B, upper panel). As in previous experiments, when RRL was supplemented with TR buffer, or when P1SHC* mRNA was used for translation, only the band corresponding to the unprocessed polypeptide was detected. When RRL was supplemented with BYL a band corresponding to the approximate size of P1 was also observed. Testing of the fractions obtained from the anion exchange chromatography showed the presence of a strong P1 band in samples complemented with the eluates of 350 mM and 400 mM KAc (Q-cEF 350 and Q-cEF 400, respectively) indicating that these fractions contained sufficient amount of factor(s) to support P1 proteolytic processing (Figure III.21B, lower panel).

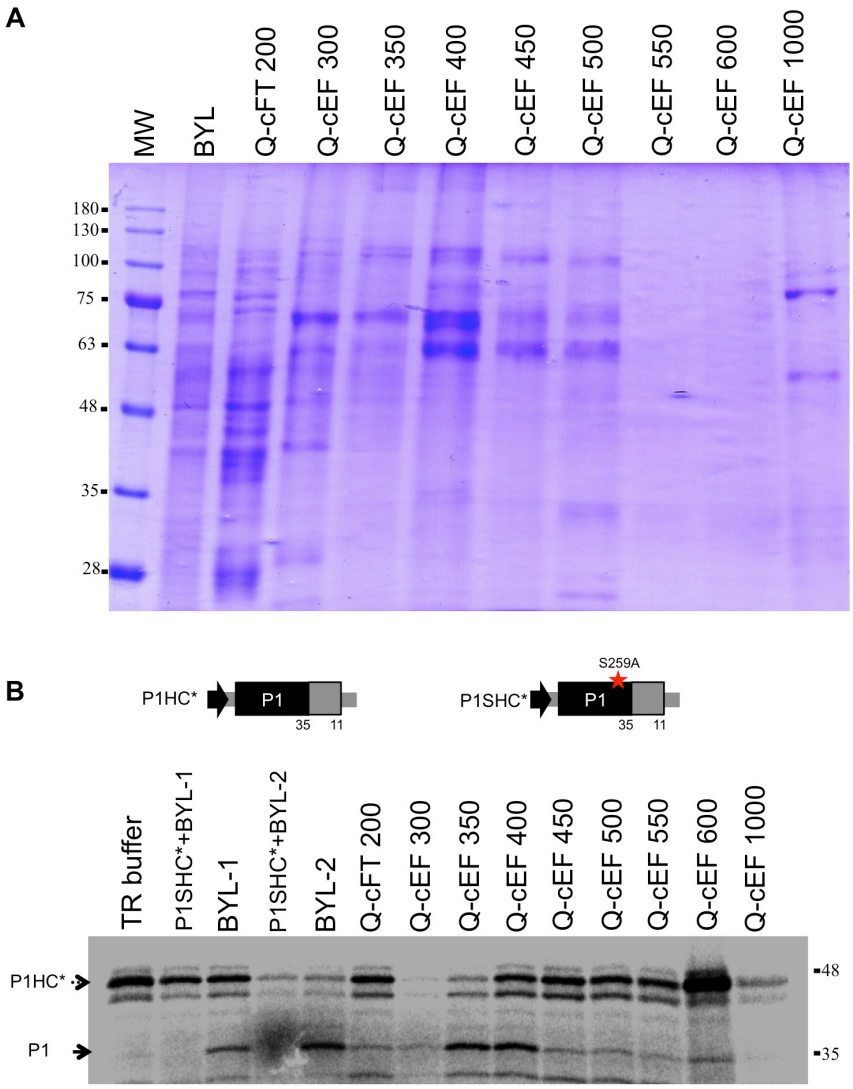

Figure III.21. Anion exchange chromatography and host factor activity assay. (**A**) 10% SDS-PAGE analysis of the different concentrated fractions obtained after anion exchange chromatography. The size of molecular weight markers run in the first lane (MW) are indicated beside the gel. (**B**) P1 proteolytic activity assay in RRL. Diagrams of DNA constructs of the coding sequences of leader proteases of PPV, amplified by PCR and subjected to *in vitro* transcription and further *in vitro* translation, T7 RNA polymerase promoter (black arrow) drives the PPV 5'UTR and the coding sequence of the P1 protease (black box) and 97 N-terminal amino acids from HC (HC*, grey box) followed by a stop codon and PPV 3'UTR. Mutation of the serine of the catalytic center (S259A) is marked with a red star. Molecular weights of the expected translation products are displayed below (upper panel). Self-cleavage activity was evaluated in RRL *in vitro* translation system with the indicated supplements. ^{35}S-labeled translation products were resolved by SDS-PAGE and detected by autoradiography. Processed and unprocessed products are marked (left); right, molecular weight markers.

To further purify the factor(s) from the complex mix of proteins still present in the eluted fractions from the Q column, gel filtration was performed using a commercial Superdex 200 column (Amersham) attached to an FPLC device. Fractions Q-cEF 350 and Q-cEF 400 were

combined (IEX-c-mix) and sample was injected into the column previously equilibrated with TR buffer (80 mM KAc). Size-based eluted fractions were collected, concentrated and analyzed by SDS-PAGE (Figure III.22A) prior to *in vitro* complementation test (Figure III.22B). Fractions 13 and 14 (GF-cEF 13 and GF-cEF 14) presented a band matching the size of P1 with the highest intensity. Proteins in these two fractions were estimated to be ~110 kDa-170 kDa in size based on column calibration. Protein complexity of the fractions was too high to perform analyses of the whole sample, but detailed comparison by SDS-PAGE of fractions GF-cEF 12 to 15 (Figure III.22C) allowed the identification of unique bands that could be responsible for the differences observed in P1 processing complementation during RRL translation (numbered 1-8). These bands were extracted from the gel and analyzed by mass spectrometry.

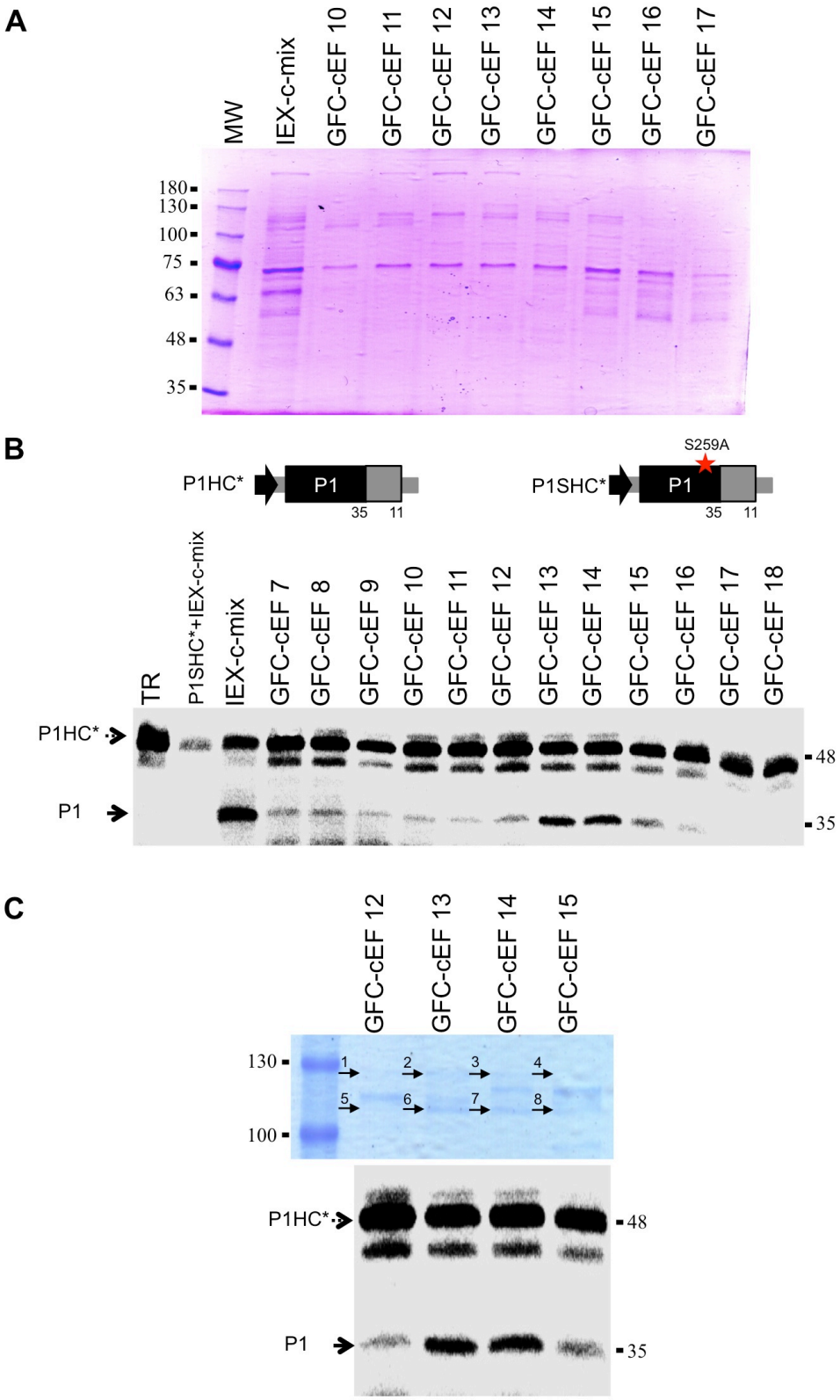

Figure III.22. Gel filtration chromatography and P1 protease activity assay. (**A**) 10% SDS-PAGE analysis of the different concentrated fractions obtained after gel filtration chromatography. The size of molecular weight markers run in the first lane (MW) are indicated beside the gel. (**B**) Diagrams of DNA constructs of the coding sequences of leader proteases of PPV, amplified by PCR and subjected to *in vitro* transcription and further *in vitro* translation, T7 RNA polymerase promoter (black arrow) drives the PPV 5'UTR and the coding sequence of the P1 protease (black box) and 97 N-terminal amino acids from HC (HC*, grey box) followed by a stop codon and PPV

3'UTR. Mutation in the serine of the catalytic center (S259A) is marked with a red star. Molecular weights of the expected translation products are displayed below (upper panel). Self-cleavage activity was evaluated in RRL *in vitro* translation system with the indicated supplements. ^{35}S-labeled translation products were resolved by SDS-PAGE and detected by autoradiography. (**C**) Detailed analysis of fractions by SDS-PAGE (upper panel) with the corresponding P1 *in vitro* activity results (lower panel). Arrows and numbers mark bands extracted from the gel and analyzed by mass spectrometry. In B and C, processed and unprocessed products (left) and molecular weight markers (right) are signaled.

III.3.2.2 Tandem anion and cation exchange chromatography previous to size exclusion purification

In order to improve plant factor isolation from BYL extract, a second P1 factor(s) purification trial was attempted including an additional cation exchange chromatographic step between anion exchange and gel filtration columns. To increase eluate sample concentration, a home-made column of ~300 µl Q resin (Amersham) was mounted. Column was equilibrated with TR buffer with 200 mM KAc. BYL extract, prepared new for the second isolation trial of the plant P1 co-factor(s), was applied to the column (11.5 mg of total protein). A two-step elution was performed using first, TR buffer containing 300 mM KAc to remove contaminants and using then, TR buffer with 500 mM KAc to ensure that P1 co-factor(s) was collected in a single fraction. Eluate was concentrated and after separating a sample for *in vitro* testing, the rest was diluted ten times with TR buffer without KAc to get a final KAc concentration of the sample prior to loading into cation exchange column of 50 mM (Figure III.23A).

Cation exchange resin (F) was packaged in a home-made column (~300µl volume) equilibrated with TR buffer with 50 mM KAc. Sample (diluted Q-cEF 500) was applied and elution was performed as in previous chromatography using a step gradient with TR buffer with increasing amounts of KAc. Collected fractions were concentrated and protein complexity of each sample was analyzed by SDS-PAGE (Figure III.23A, upper panel). Fractions were then tested for P1 cleavage complementation in RRL fed with P1HC* transcript. In this case, the fractions that presented the highest complementing activity were those eluted with 300 mM KAc and 400 mM KAc (F-cEF 300 and F-cEF 400) (Figure III.23A, bottom panel). Fraction F-cEF 300 was used to continue with size exclusion chromatography in a Superdex 200 column as before. Eluted fractions were concentrated and protein complexity was analyzed by SDS-PAGE. Silver staining was used after unsuccessful band detection with Coomasie staining (Figure III.23B, upper panel). Fraction GF-cEF 14, showing the highest P1 processing complementation was used for further analysis by mass spectrometry (Figure III.23B, bottom panel).

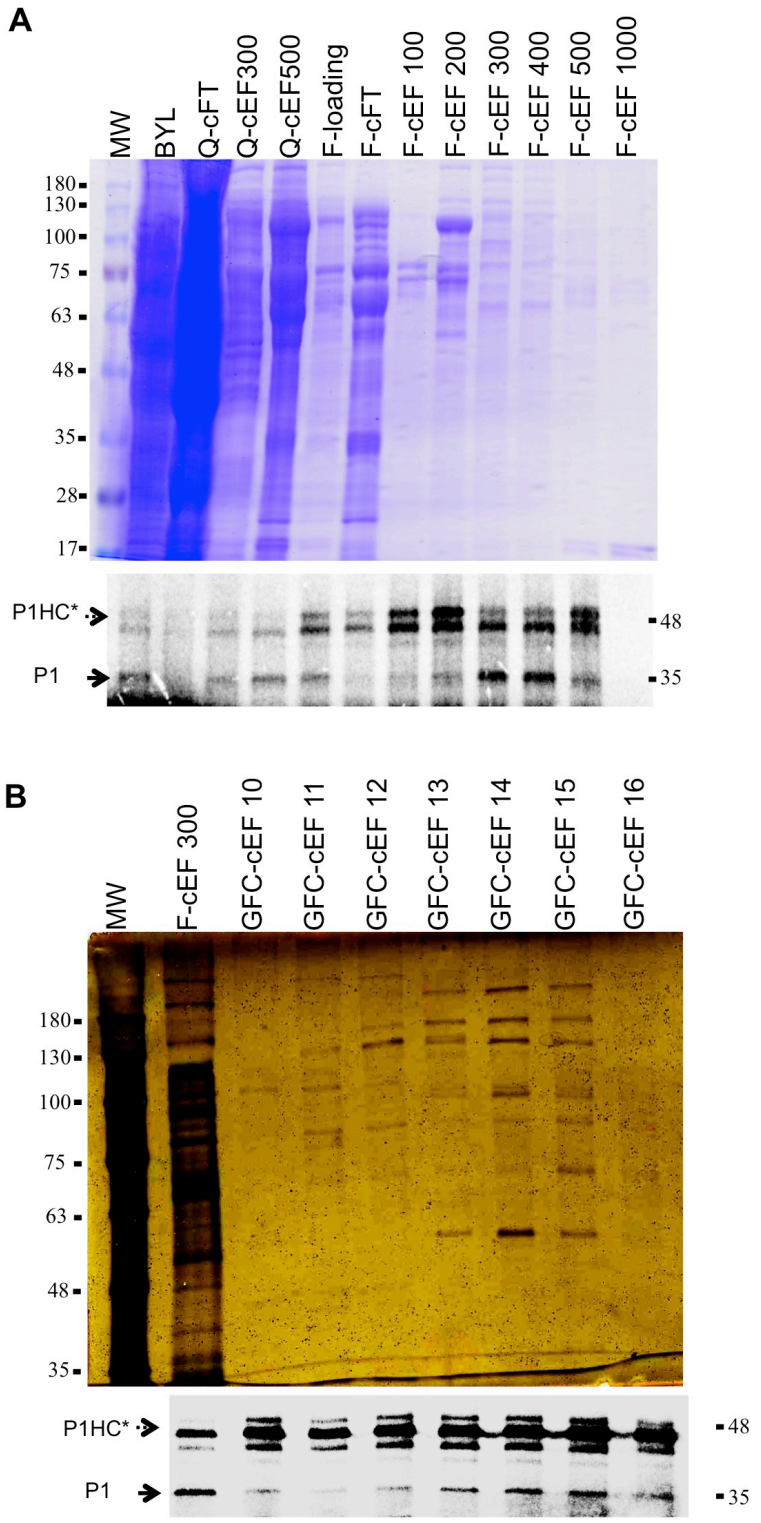

Figure III.23. Second P1 co-factor(s) purification trial: ion exchange, gel filtration chromatography and P1 protease activity assay. (**A**) 12% SDS-PAGE analysis of the different concentrated fractions obtained after cation exchange chromatography (upper panel) followed by P1 proteolytic activity assay in RRL (lower panel). (**B**) Silver stained 12% SDS-PAGE analysis of the different concentrated fractions obtained after gel filtration chromatography (upper panel) followed by P1 proteolytic activity assay in RRL (lower panel). In upper panels, the size of molecular weight markers run in the first lane (MW) are indicated beside the gel. In lower panels, processed and unprocessed products (left) and molecular weight markers (right) are indicated.

A third purification trial of the P1 complementation factor(s) was performed following the new anion exchange- cation exchange- gel filtration chromatographic pipeline. New BYL extracts were prepared. In order to increase plant factor(s) yield, BYL extract initial amount was increased (14.4 mg of total protein) and instead of single ~300 µl Q and F columns, three independent home-made columns from each resin were used in parallel. Same fractions as in the second trial were collected from Q purification step (Q-cEF 500 from each column). After cation exchange, fractions helping to produce the highest accumulation of processed P1 in the RRL complementation assay were F-cEF 200 and 300 (Figure III.24A). These fractions were combined and used to perform size exclusion chromatography as before. After gel filtration, protein complexity of the eluted fractions was analyzed by SDS-PAGE followed by silver staining. Concentrated fractions were used for *in vitro* complementation analyses and two of these fractions were analyzed by mass spectrometry: fraction GF-cEF 11, which presented little P1 processing complementation and could be used as negative control, and fraction GF-cEF 14, which showed the highest P1 proteolytic complementation activity in RRL (Figure III.24B).

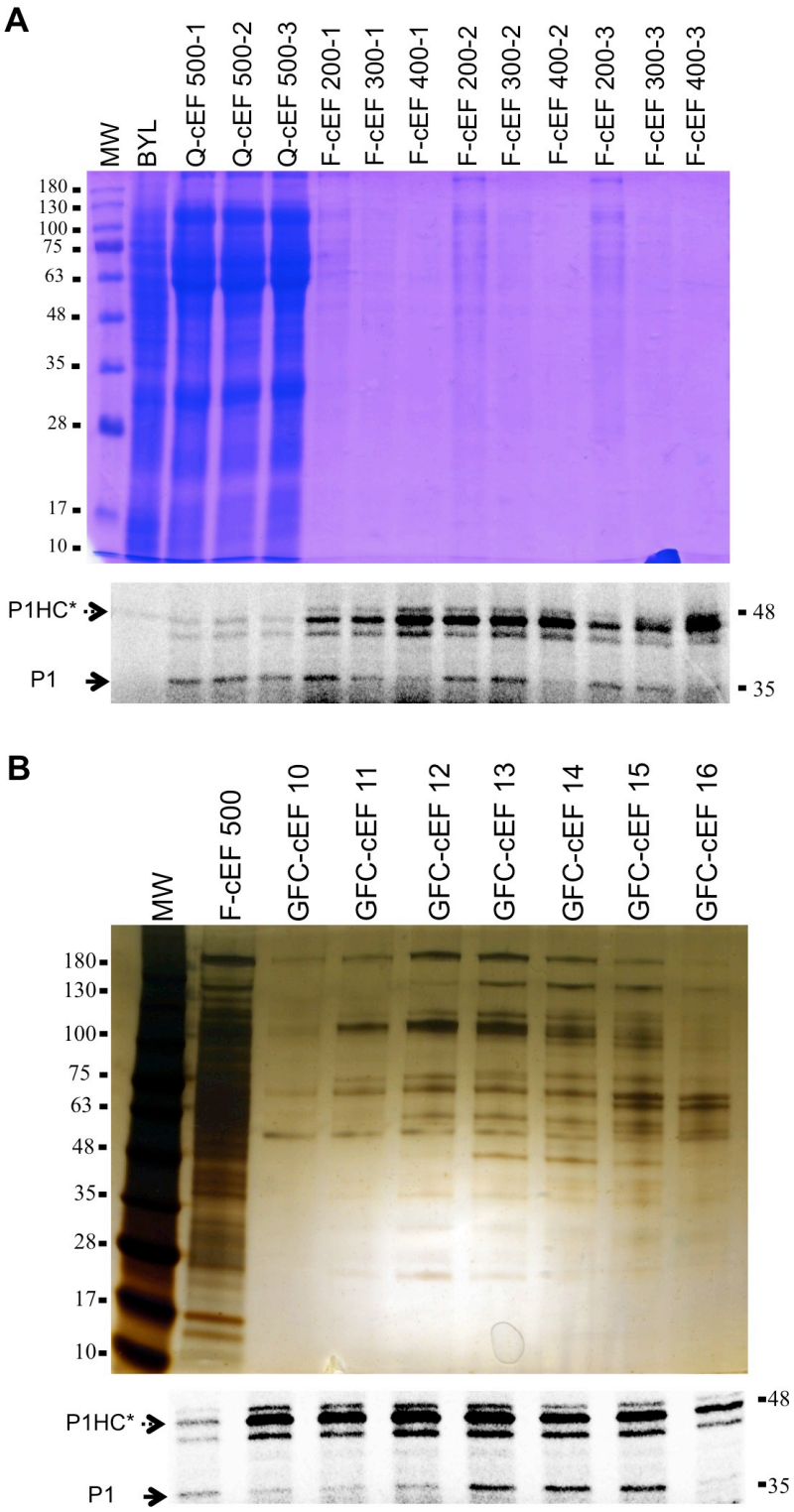

Figure III.24. Third P1 co-factor(s) purification trial: ion exchange, gel filtration chromatography and P1 protease activity assay. (**A**) 12% SDS-PAGE analysis of the different concentrated fractions obtained after ion exchange chromatography (upper panel) followed by P1 proteolytic activity assay in RRL (lower panel). (**B**) Silver stained 12% SDS-PAGE analysis of the different concentrated fractions obtained after gel filtration chromatography (upper panel) followed by P1 proteolytic activity assay in RRL (lower panel). In upper panels, the size of molecular weight markers run in the first lane (MW) are indicated beside the gel. In lower panels, processed and unprocessed products (left) and molecular weight markers (right) are indicated.

III.3.3 Identification of proteins in purified samples and selection of P1 plant co-factor(s) candidates

III.3.3.1. Mass spectrometry analyses

From the initial purification trial, differences in protein band pattern could be observed between fractions GF-cEF 12-15 (Figure III.22C). Mismatching bands were selected to be extracted from the gel and specifically identify proteins present in those bands by MALDI-TOF analysis. Identified peptides were confronted with a *N. tabacum* database (https://solgenomics.net) resulting in the identification of six proteins listed in Table III.1 (annex 1). Fractions GF-cEF 13 and 14 showed in the *in vitro* complementation test (Figure III.22) the highest accumulation of a band corresponding to the processed P1. This suggests that these fractions should contain the highest P1 co-factor(s) accumulation. Based on the number of peptides identified, two proteins from the list were assumed to be present in higher amounts in GF-cEF 13-14 than in the other two fractions: Ubiquitin activating enzyme 2 (46936) and TUDOR-SN protein 1 (85936). Ubiquitin specific protein 12 (8285) (USP12) and tRNA synthetase class I (88669) were present in GF-cEF 13-14, but peptides were identified in other fractions as well. The other two proteins, chaperones of the hsp70 family, were practically absent from GF-cEF14.

Table III.1. PPV P1 proteolytic host factor indentification by MALDI-TOF analysis.

	Fraction 12	**Fraction 13**	**Fraction 14**	**Fraction 15**
P1 processing	Δ	Δ Δ Δ Δ Δ	Δ Δ Δ Δ Δ	Δ Δ
Analyzed bands	**1**	**2**	**3**	**4**
Ubiquitin specific protein 12 (8285)	+++	++++	++	+
Ubiquitin activating enzyme 2 (46936)	++	+++++	+++++	++++
Analyzed bands	**5**	**6**	**7**	**8**
Heat shock protein 70 (HSP70) (103985)	++++	+	-	-
Heat shock protein 70 (HSP70) (119001)	+++++	+++	+	-
TUDOR-SN protein I (85936)	+	+++++	+++++++	+++
tRNA synthetase class I (88669)	-	++	++	+++
Ubiquitin activating enzyme 2 (46936)	-	-	-	+

Δ: P1 cleavage efficiency according to Fig. III.22.
-/+: Protein presence based on number of peptides identified.

In order to get a complete set of possible factors that might be involved in P1 proteolytic processing, whole fraction GF-cEF 14 from the second purification trial (Figure III.23) was used for protein identification by liquid chromatography coupled to electrospray tandem mass

spectrometry (LC ESI-MS/MS). Same database as before was used to match the corresponding identified peptides and a list of 33 proteins was obtained (Table III.2, annex 2). TUDOR-SN 1 protein (85936) previously identified was also detected in this second analysis with the highest number of peptides. Two other isoforms of this protein were identified (56422 and 52854). No peptides corresponding to Ubiquitin activating enzyme 2 (46936) were detected, indicating that this protein was probably isolated out of the extract during the cation exchange chromatography added to the purification scheme. On the other hand, Ubiquitin specific protein 12 (8285), previously identified was present in the whole sample analysis together with two other isoforms (45188 and 54907). Similarly, two chaperons of the hsp70 family were detected, but none of them corresponded to the same exact protein as the ones previously detected. All identified proteins are listed in Table III.2 (annex 2).

The same LC ESI-MS/MS procedure was used to analyze proteins obtained from the last purification trial. In this case, two fractions GF-cEF 11, with practically no presence of P1 band in the *in vitro* analysis, and GF-cEF 14, with the strongest signal corresponding to the processed P1, were used (figure III.24). 40 and 77 proteins, respectively, were identified. Details can be found in Table III.2 (annex 3).

Table III.2. Protein from the second and third purification trials identified after LC-MS/MS(ESI) analysis

	Fraction 14 (Exp. 2) protein name (mRNA cds*)	**Fraction 14 (Exp. 3) protein name (mRNA)**	**Fraction 11 (Exp. 3) protein name (mRNA)**
P1 processing	Δ Δ Δ Δ Δ	Δ Δ Δ Δ Δ	-
1	TUDOR-SN protein 1 (85936, 56422, 52854)	TUDOR-SN protein 1 (85936, 56422, 52854, 130728, 52858)	
2	Chorismate synthase (3181)	Chorismate synthase (3182)	
3	Ubiquitin-associated (UBA) / TS-N domain-containing protein (120910)	Ubiquitin-associated (UBA)/TS-N domain-containing protein (120910, 127871)	
4	NADH-dependent glutamate synthase 1 (22014, 102188, 29629)	NADH-dependent glutamate synthase 1 (22014, 102188, 29629)	NADH-dependent glutamate synthase 1 (22013, 102188, 29629)
5	La protein 1 (33365)	La protein 1 (33365)	La protein 1 (33365)
6	Heat shock protein 70 (Hsp 70) family protein (103984, 56510)	Heat shock protein 70 (Hsp 70) family protein (119001, 2408, 94378, 103985, 56510, 15738, 12846)	Heat shock protein 70 (Hsp 70) family protein (119001, 94378, 103985, 56510, 15738)
7	Ubiquitin-specific protease 12 (8285, 45188, 54097)	Ubiquitin-specific protease 12 (8285, 6538, 54097,90497)	
8	Heat shock protein 89.1 (20701)	Heat shock protein 89.1 (20701)	

9	Lysyl-tRNA synthetase 1 (38895)	tRNA synthetase class I (I, L, M and V) family protein (88668)	
10	Unknown protein (28381, 2683)	Unknown protein (28381,109413, 111386, 107854, 121620)	
11	Transducin/WD40 repeat-like superfamily protein (65900)	Transducin/WD40 repeat-like superfamily protein (92935)	Transducin/WD40 repeat-like superfamily protein (92935, 64331, 46708)
12	DEA(D/H)-box RNA helicase family protein (32753)		
13	L-Aspartase-like family protein (46598)		
14	Eukaryotic translation initiation factor 2 gamma subunit (11051), beta subunit (24094)	Eukaryotic translation initiation factor 2 subunit 1 (82151)	Eukaryotic translation initiation factor 3 A (126735, 111605), 3B1 (66895), 3C (29398), 3E (13095), 3G1 (15484), 3H1 (8216)
15	Glycine decarboxylase P-protein 2 (33746)		
16	Nitrogen fixation S (NIFS)-like 1 (4210)		
17	Chromatin remodeling 42 (27829)		
18	Nodulin MtN3 family protein (27829)		
19	D-isomer specific 2-hydroxyacid dehydrogenase family protein (86325)		
20	Tryptophan synthase beta type 2 (4649)		
21	Ucleoside diphosphate kinase family protein (48876)		
22	Forkhead-associated (FHA) domain-containing protein (134800)		
23	ARM repeat superfamily protein (81166)	ARM repeat superfamily protein (114876)	
24	Global transcription factor group E8 (58089)		Global transcription factor group E8 (58089)
25		Heat shock protein 101 (107329, 46625)	
26		Heat shock protein 91 (117463)	Heat shock protein 91 (117463)
27		Glycine-rich protein (10661, 114497, 45863)	
28		Isovaleryl-CoA-dehydrogenase (105558)	
29		Rotamase FKBP 1 (6740)	
30		S-adenosyl-L-methionine-dependent methyltransferases superfamily protein (85755, 85754)	
31		2-oxoglutarate dehydrogenase, E1 component (111073)	

32		CAP-binding protein 20 (48309)	
33		Ribosomal RNA processing 4 (15358)	Ribosomal RNA processing 4 (15358)
34		Cleavage and polyadenylation specificity factor (CPSF) A subunit protein (47529, 130628)	
35		Nucleic acid-binding, OB-fold-like protein (44788)	Nucleic acid-binding, OB-fold-like protein (44788, 50328)
36		SUMO-activating enzyme 2 (16050)	
37		Ribosomal protein S5 domain 2-like superfamily protein (117612)	Ribosomal protein S5 domain 2-like superfamily protein (117612, 3528), S6 family protein (50919)
38		Pyruvate dehydrogenase E1 alpha (23594)	
39		DEK domain-containing chromatin associated protein (64643)	
40		Translin family protein(94536)	Translin family protein (94537, 16151)
41		3'-5'-exoribonuclease family protein (91804, 18720, 119444)	3'-5'-exoribonuclease family protein (91804, 18721, 119444)
42		Chaperone protein htpG family protein (51789)	
43		PNAS-3 related (108808)	PNAS-3 related (108808)
44		Adaptin family protein (94729)	Adaptin family protein (94729)
45		Phosphoenolpyruvate carboxykinase 1 (77874)	Phosphoenolpyruvate carboxykinase 1 (77874)
46		Small nuclear ribonucleoprotein family protein (76703)	
47		Protein kinase superfamily protein (25528)	
48		Methyl-CPG-binding domain 11 (33535)	
49S		Nucleolin like 2 (58161)	
50		ATP binding (104145)	
51		Cytochrome P450 superfamily protein (7897)	
52		Mitochondrial lipoamide dehydrogenase 1 (21974)	
53		Cullin-associated and neddylation dissociated (982)	Cullin-associated and neddylation dissociated (41849)
54		Mediator complex, subunit Med7 (19486)	
55		Actin 7 (9884)	
56		Minichromosome maintenance (MCM2/3/5) family protein (45834)	Minichromosome maintenance (MCM2/3/5) family protein (29344, 6439, 92935, 45834)
57		Rhamnose biosynthesis 1 (247)	Rhamnose biosynthesis 1 (247)
58		Methyltransferase 1 (82133)	
59		Putative adipose-regulatory protein	

		(Seipin) (66203)	
60		Heat-shock protein 70T-2 (59674)	Heat shock cognate protein 70-1 (12925)
61		Heat shock protein 90.1 (35038)	
62			DegP protease 7 (51109, 18273, 131633)
63			Arginine methyltransferase 11 (16984)
64			Ubiquitin-protein ligase Cullin 4 (106876)
65			GTP binding Elongation factor Tu family protein (18823)
66			Cullin 1 (28644)
67			Phenylalanyl-tRNA synthetase, putative / phenylalanine--tRNA ligase (58452)
68			Splicing factor-related (15215)
69			NADP-malic enzyme 4 (52564)
70			UBX domain-containing protein (30772)
71			2-cysteine peroxiredoxin B (46583)
72			Co-chaperone GrpE family protein (7594)
73			COP9 signalosome, subunit CSN8 (18633)
74			Cysteine-rich RLK (RECEPTOR-like protein kinase) 8 (58805)
75			General regulatory factor 8 (23462)
76			ATPase, V1 complex, subunit B protein (25797)
77			Arginosuccinate synthase family (36775)

Δ/-: P1 cleavage efficiency according to Figures III.23 and III.24.
*mRNA cds: Accesion numbers as indicated in Sol Genomics Network (https://solgenomics.net/).

III.3.3.2 Selection of candidates

Theoretically, the plant factor(s) involved in P1 processing complementation during RRL translation should be among the list of 33 proteins obtained from the mass spectrometry analysis of the fraction GF-cEF 14 from the second isolation experiment and also should be included in the list of 77 proteins detected in fraction GF-cEF 14 of the third attempt of extract purification (Table III.2). Thus, information from both lists can be cross-referenced to obtain a list of 13 common proteins, which necessarily should include the unidentified P1 co-factor(s). Considering that proteins present in fraction GF-cEF 11 of the third purification trial most likely were not involved in P1 proteolytic activity, the list of possible candidates could be reduced from 13 to 8 proteins, subtracting the candidates common to both cases (shadowed in grey in Table III.2).

Given the disordered nature of the N-terminal part of P1, a chaperone fits very well the anticipated function of the plant co-factor. Thus, the list of eight candidates was completed including chaperones from the hsp70 family previously identified (in the first band analysis and in the second and third whole sample analyses). Additionally, other features were considered to select among this list of candidates the most promising ones to be further validated or at least to rank the selected candidates in order of likelihood (Table III.3): i) mass spectrometry reliability: number of peptides identified and quality of identification; ii) function: having a role than can be related to P1 processing such as the one of chaperones; iii) presence in the first mass spectrometry analysis in the fractions showing P1 proteolytic complementation; iv) size of the protein: having the expected size based on the last purification step; v) reported as possible P1 interactor (Martínez & Daròs, 2014).

Table III.3 Candidates selected for validation in order of likelihood.

Candidates	Selection criteria of candidates
Tudor-SN protein 1	Highest identification score of all candidates Identified as factor in the first band analysis Protein size fit the criteria (108 kDa) Reported as P1 interactor (Martínez & Daròs, 2014).
Heat shock protein 89.1	Medium identification score Chaperone function related with protein folding Protein size fit the criteria (90 kDa)
Ubiquitin-specific protease 12	High identification score Protein size fit the criteria (131 kDa)
Heatshock protein 70 (Hsp 70) family protein	High identification score Chaperone function related with protein folding Identified as non-factor in the first band analysis Protein size does not fit the criteria (70KDa) Reported as P1 interactor (Martínez & Daròs, 2014).
Ubiquitin-associated (UBA) / TS-N domain-containing protein	High identification score Protein size fit the criteria if acting as a dimer (47 kDa)
unknown protein	Low identification score Unknown function Protein size does not fit the criteria (36 kDa)
Chorismate synthase	High identification score Protein size fit the criteria if acting as a dimer (50 kDa)
tRNA synthetase class I (I, L, M and V) family protein	Low identification score Identified as non-factor in the first band analysis Protein size fit the criteria if acting as a dimer (62 kDa)
ARM repeat superfamily protein	Very low identification score Protein size fit the criteria (99 kDa)

III.3.4. Validation of host factor candidates involved in P1 proteolytic autocleavage

III.3.4.1 Use of protein inhibitors to abolish *in vitro* RRL complementation of P1 self-cleavage

Function inhibition of the selected plant factor(s) present in the BYL extract might hamper, partially or totally, P1 proteolytic processing complementation. From the list of eight factors selected for validation (Table III.4), functional inhibitors were found for the chaperone activity of hsp70 and hsp90 protein families (Pifithrin μ and 17AA, respectively) and for the nuclease activity of TUDOR-SN 1 protein (deoxythymidine 3', 5'-bisphosphate, pdTp). Chaperone's inhibitors were only soluble in ethanol and this compound disrupts *in vitro* protein translation, making the following tests unreliable. pdTp, on the other hand, was soluble in water and it was tested for its ability to abolish P1 protease complementation by BYL purified extract.

P1HC* construct was used together with P1ProHC*. This last construct, previously tested (Figure III.10, 11), was included as a negative control because it expresses a P1 protease domain that acts independently of a host factor and should not be affected by pdTp (Figure III.25A, C). Triplicates were used to perform quantification and statistical analysis of the P1 processing after inhibitory treatment (Figure III.25B, D). Purified BYL sample (F-cEF 300) was used to supplement RRL and complement P1 processing during translation. A band matching the size of P1 could be detected in all cases, but adding increasing amounts of inhibitor occasioned a significant reduction of this band and P1 processing decreased from 30% (without pdTp) to 14% (with 2.5 mM pdTp). This result suggested a possible negative effect on P1 processing complementation driven by the TUDOR nuclease inhibitor. However, same experiment performed with the unrestricted protease P1Pro released similar results (Figure III.25 C and D) with protease catalytic activity decreasing from 39% (no inhibitor) to 24% (with 2.5 mM pdTp). Experiment was repeated using P1HC* and P1ProHC* without pdTp and with 2.5 mM pdTp and similar results were obtained. This indicates that the observed effect is not linked to a reduction of the complementation activity of the P1 protease co-factor.

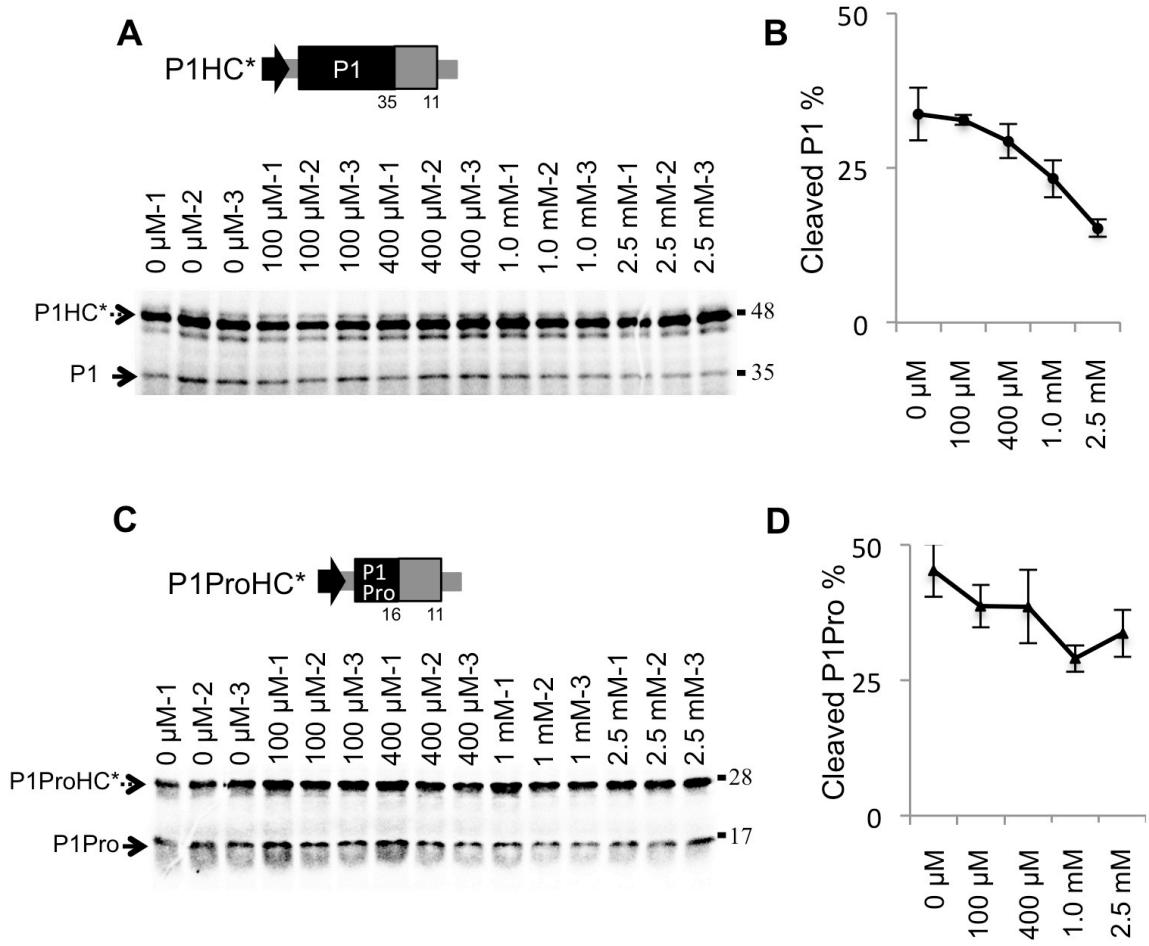

Figure III.25. Tudor-SN protein inhibition assay. (**A,C**) Diagrams of DNA constructs of the coding sequences of leader proteases of PPV, amplified by PCR and subjected to *in vitro* transcription and further *in vitro* translation, T7 RNA polymerase promoter (black arrow) drives the PPV 5'UTR and the coding sequence of the P1 protease (black box) and 97 N-terminal amino acids from HC (HC*, grey box) followed by a stop codon and PPV 3'UTR. Molecular weights of the expected translation products are displayed below (upper panels). Self-cleavage activity was evaluated in RRL *in vitro* translation system with the corresponding supplements and increasing amounts of pdTp. ^{35}S-labeled translation products were resolved by SDS-PAGE and detected by autoradiography. Processed and unprocessed products are marked (left); right, molecular weight markers (lower panels). (**B, D**) Quantification analysis of P1 and P1Pro processing efficiency according to the *in vitro* translation assay (A) and (C). Plots were built based on P1 cleavage efficiency estimated by the program Quantity One. One hundred was assigned to maximum translation value, which includes processed and unprocessed products of P1 construct.

III.3.4.2 Use of VIGS for candidate validation in *N. benthamiana* plants

Similar to the approach followed by the use of inhibitors under *in vitro* translation conditions, VIGS was used to downregulate the expression of the corresponding candidates in *Nicotiana* plants and observe the effect of this gene silencing on P1 processing ability. *N. benthamiana* genes presenting >98% identity to the candidates found in *N. tabacum* were identified by BLAST (https://solgenomics.net) and selected for silencing. TRV was used as VIGS vector. The first six candidates from Table III.3 were selected for the initial test. pTRV2 carrying a plant-unrelated sequence from GUS (pTRV2-GUS) was used as negative control. Another pTRV2 plasmid carrying a sequence to silence phytoene desaturase gene expression (pTRV2-PDS) was used as control of TRV infection (Figure III.25A). *N. benthamiana* plants were co-agroinoculated with pTRV1 and the corresponding modified pTRV2 plasmid (or plasmids) to silence the candidate genes (details in sections II.5.1 and II.16). Phenotype development was recorded (Table III.4). After 12 days a clear phenotype was starting to show in some of the silenced plants compared to pTRV2-GUS inoculated plants. The most dramatic effect was observed in plants inoculated with pTRV2-HSP70, which died after 16-18 days. Other plants, such as the ones inoculated with pTRV2-HSP89 and pTRV2-USP12, also showed phenotypic variations with respect to pTRV2-GUS-inoculated plants. pTRV2-UBA-, pTRV2-Unknown- and pTRV2-TUDOR-inoculated plants did not present any phenotypic variation. The latter plants were selected for further analysis of candidate gene expression. Tissue was collected at 12 dpi and RT-qPCR analysis confirmed an 80% reduction in Tudor-SN 1 gene expression in these plants compared to the control pTRV2-GUS-inoculated plants.

Table III.4. Symptomatology of plants inoculated with modified TRV vectors

	Phenotype at 12 dpi
pTRV1+pTRV2-GUS	Death of the inoculated leaves and chlorosis in the upper non-inoculated leaves.
pTRV1+pTRV2-PDS	Death of the inoculated leaves and general dwarfing of the plant; Upper non-inoculated leaves show strong whitening.
pTRV1+pTRV2-TUDOR	Similar to pTRV1+pTRV2-GUS inoculated plants.
pTRV1+pTRV2-HSP89.1	Death of the inoculated leaves and general dwarfing of the plant; Upper non-inoculated leaves show strong necrotic symptoms and youngest leaves die after 17-20 dpi.
pTRV1+pTRV2-HSP70	Death of the inoculated leaves and general dwarfing of the plant Chlorosis in the upper non-inoculated leaves until 5-7 dpi when plant start to gradually die. Plants are dead by 16-18 dpi.
pTRV1+pTRV2-USP12	Death of the inoculated leaves Reduction in the number of new leaves that grow bigger, darker and thicker. At 15-17 dpi, non-inoculated leaves become fragile, easily falling off the plant.
pTRV1+pTRV2-UBA	Similar to pTRV1+pTRV2-GUS inoculated plants
pTRV1+pTRV2-Unknown	Similar to pTRV1+pTRV2-GUS inoculated plants

Initially, P1HC-TAP was used as reporter of P1 processing. This construct consists of P1HC from PPV followed by a TAP tag at the C-terminal part of HC to allow optimal identification of this protein by immunoblot assays (Figure III.26A). Based on our previous observations, *N. benthamiana* plants were inoculated with the different TRV constructs and after 12 days, when pTRV2-PDS-inoculated plants showed strong whitening of the first two non-inoculated leaves, corresponding leaves were agroinfiltrated in the other plants with plasmid carrying P1HC-TAP. pTRV2-HSP70-inoculated plants were treated with P1HC-TAP 5 days earlier to be able to recover some data before the death of the plants.

Two days after the second agroinoculation, leaves were collected and a TAP-specific western blot analysis of the infiltrated tissue was performed. Processed and unprocessed P1HC-TAP products were analyzed comparing expression in plants treated with pTRV2-candidate gene- and control pTRV2-GUS-inoculated plants. A reduction of the P1-complementing factor in *N. benthamiana* originated by VIGS might cause a decrease in P1 processing. Due to the unstable nature of P1HC (Figure III.6), an increase in this unprocessed product is not anticipated. However, a decrease in the processed HC is plausible and would be a strong indication that P1 proteolytic activity is diminished. A band corresponding to P1HC-TAP can be observed in samples presenting high protein expression, but as anticipated, there is no correlation with VIGS silencing. On the other hand, a clear reduction in HC-TAP accumulation was observed in the case of pTRV2-HSP70 and some HC-TAP reduction could be detected in pTRV2-USP12-

inoculated plants compared to pTRV2-GUS control plants. Plants treated with pTRV2-TUDOR, pTRV2-HSP89, pTRV2-Unknown protein and pTRV2-UBA did not show significant differences in HC-TAP accumulation (Figure III.26 B).

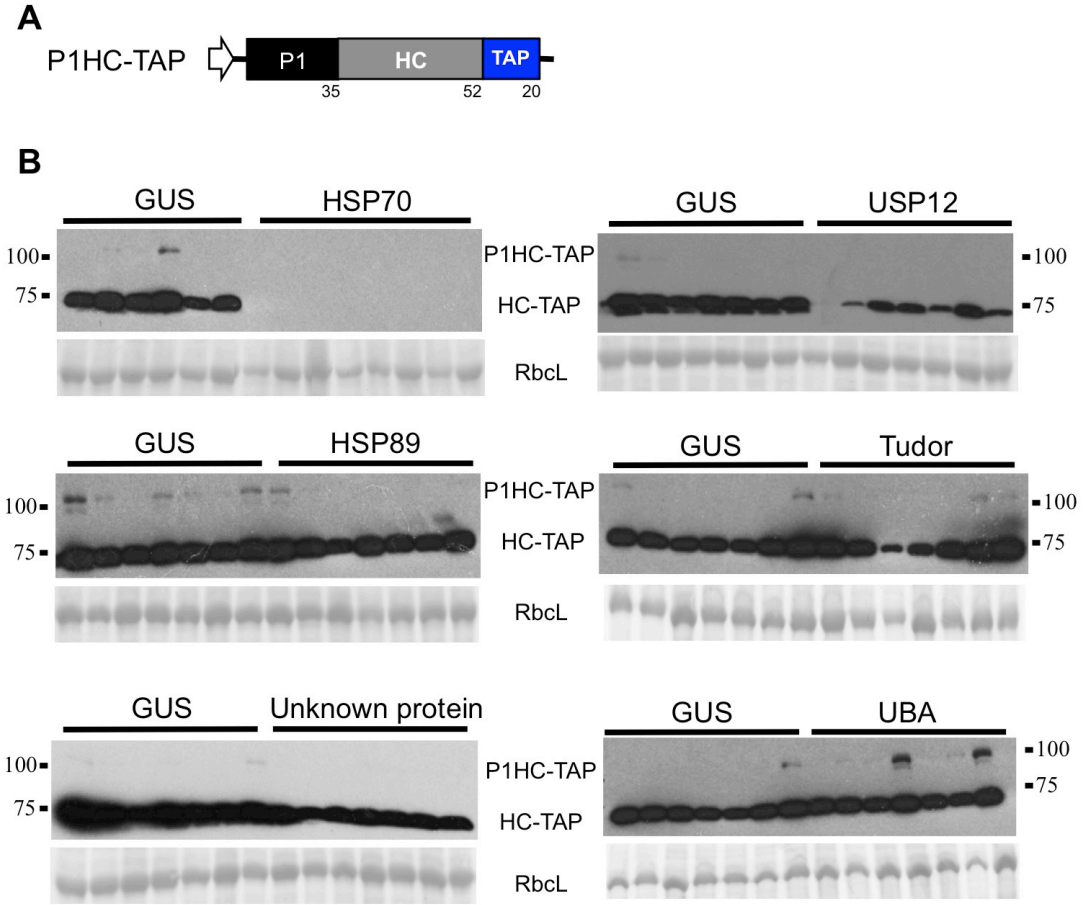

Figure III.26. Candidates VIGS test in *N. benthamiana*. (**A**) Diagram of the transient expression construct agroinfiltrated in *N. benthamiana*. Molecular weight of the protein products is shown below. (**B**) TAP-specific immunoblot analysis performed two days after agroinfiltration of the transient expression constructs. pTRV2 inoculated constructs are indicated by the silenced gene. Processed and unprocessed products are marked between the panels; outside, molecular weight markers. Each lane corresponds to a single plant. Ponceau red-stained blots showing RbcL were included as loading controls.

Considering the strong phenotype that pTRV2-HSP70- and pTRV2-USP12-inoculated plants displayed, the observed HC-TAP reduction could be explained by deficient protein expression in these plants unrelated to P1 cleavage. To test this, another experiment was performed using P1HC-TAP as before, but including a construct expressing TAP-HC to control protein translation in these plants (Figure III.27.A). As in the previous experiment, two days after the inoculation of the second transient expression clones, leaves were collected and analyzed by western blot analysis. No unprocessed P1HC-TAP product could be detected this time, probably

due to reduced exposure of the membrane to the antibody. A reduction in HC-TAP processed product was observed, as before, in plants inoculated with pTRV2-HSP70 and pTRV2-USP12 compared to the pTRV2-GUS control inoculated plants. This reduction, however, is also observed when TAP-HC construct was employed indicating that at least part of the effect was caused by a reduction in protein translation and not by a reduction of P1 processing.

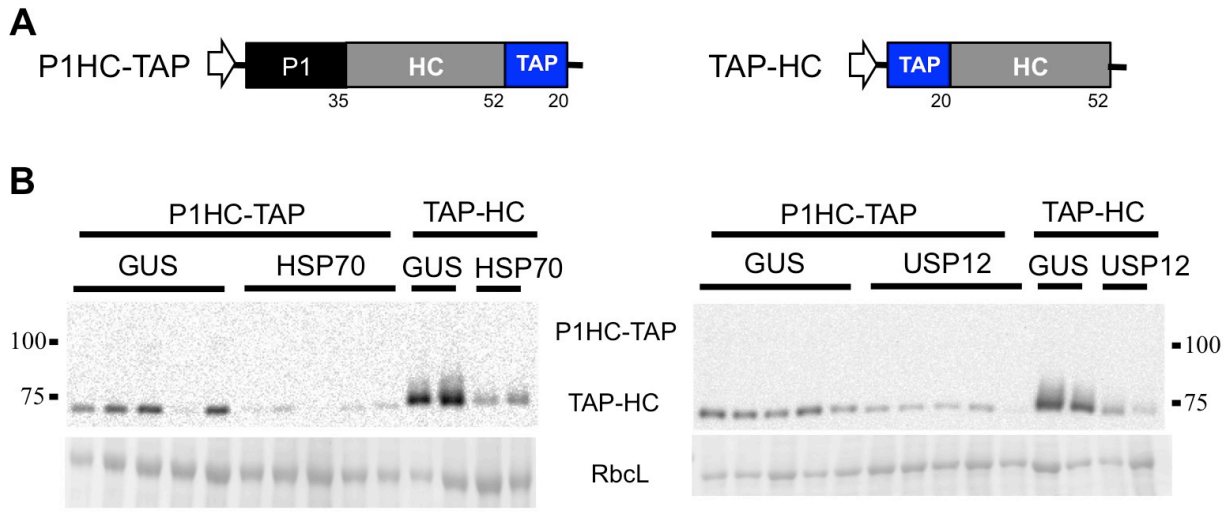

Figure III.27. Transient expression of P1HC-TAP and TAP-HC in TRV-HSP70- and TRV-USP12-inoculated plants. (**A**) Diagram of the agroinfiltrated constructs. Molecular weight of the protein products is shown. (**B**) anti-TAP inmunoblot analysis performed two days after agroinfiltration of the transient expression constructs. pTRV2 inoculated constructs are indicated by the silenced gene. Processed and unprocessed products are marked between the panels; outside, molecular weight markers. Each lane corresponds to a single plant. Ponceau red-stained blots showing RbcL were included as loading controls.

DISSCUSSION

IV. Discussion

Viral endopeptidases are proteins present in most viruses following the polyprotein strategy for the translation of their proteins. They are essential for the processing of the initial product into the final functional units, and at the same time, processing is used in many cases as a tool for the modulation of varied functions such as replication, virion formation or host range delimitation. In addition, endopeptidases play a role in the counterdefense viral mechanism and might display other features still to be defined. Being key players in the viral life cycle, in animal viruses these proteinases have been target of antiviral therapies for many years. Plant viral proteases are still behind in this regard and, with a few exceptions, current knowledge about function, structure or their use as biotechnological tools remains scarce.

This thesis focuses on leader proteases of the *Potyviridae* family as essential components during viral infection trying to characterize their mechanism of action, their involvement in host range definition and to identify the plant factor(s) implicated in their proteolytic processing.

IV.1 Type A *vs* Type B P1 proteins

P1 leader proteases are present in all potyviruses and they are thought to be, at least in part, responsible for the large expansion and host range diversification of the genus, as suggested by previous works (Rodamilans et al., 2013;Valli et al., 2007;Desbiez & Lecoq, 2004;Maliogka et al., 2012;Salvador et al., 2008b) and further supported by the data presented in this thesis. This role seems to be directly dependent on the requirements of the protein for a plant co-factor(s). Acting as a negative regulator during viral infection, PPV P1 is able to modulate the severity of infection and balance viral amplification and plant response (Pasin et al., 2014b). This function, arguably common for all P1 potyvirus orthologs and, by similarity, rymoviruses, is however absent in P1 proteins of tritimoviruses, poaceviruses (also assumed on the basis of sequence similarity), BlVY, the sole member of the genus *Brambyvirus* and in a subset of P1 proteins of ipomoviruses (Figure III.1 and IV.1).

The discovery of the tandem array of P1 leader proteases in the ipomovirus CVYV prompted to establish new terminology to define P1 proteins as P1a and P1b (Janssen et al., 2005;Valli et al., 2007). This classification was not merely semantic since important differences separated the proteins (Rodamilans et al., 2013). In this thesis different potyvirid genera were studied to define the characteristics of each leader protease, and it is proposed that Type A and Type B P1 classification was used in future works for a better understanding of the characteristics of each protein. The experiments performed clearly reinforce the mechanistic difference of Type A *vs*

Type B P1 proteins, emphasizing the requirement or not of a plant co-factor(s) to develop proper protease activity. Previously defined characteristics such as the isoelectric point distinction and the specialization of Type B P1 proteins as RNA silencing suppressors are also described (Figure III.2 and IV.1). RSS activity has been previously shown for the ipomoviral CVYV P1b (Valli et al., 2006) and CBSV P1 (Mbanzibwa et al., 2009), for tritimoviral WSMV P1 (Young et al., 2012) and for TriMV P1 of poaceviruses (Tatineni et al., 2012).

Experiments were done in which the RSS activity of the tritimoviral BStMV P1 was confirmed, whereas that activity was absent in the brambyvirus BlVY protease (Figure III.2). In this case, assay conditions may not be suitable for detecting BlVY P1 RSS, or BlVY P1 activity may depend on host-specific features. In addition, this isolate has an atypical large P1 including an AlkB domain (Susaimuthu et al., 2008) that could affect protein functionality. It should be noted that the genus *Brambyvirus* is understudied with information available for a single species. Others, yet to be discovered, species from this genus could reveal alternative sequence features helping to understand this particular protease, as it was the case with an AlkB domain-containing protein in the sister family *Secoviridae* (Halgren *et al.*, 2007;McGavin *et al.*, 2010).

A feature that both Type A and B P1 proteins have in common is their proteolytic processing exclusively *in cis*. Previous data regarding the alphavirus leader serine protease togavirin (Aggarwal et al., 2014) suggested the possibility of a functional reversion in the way P1 proteins work by deletion of C-terminal amino acids. The data presented in this thesis (Figures III.3 and III.4) indicate that P1 leader proteases cannot revert to an *in trans* behavior as togavirin does, once the final amino acid is removed. It is possible that the tested deletions were not suitable for the intended change of proteolytic mechanism, but having in mind that three different combinations were tested, it seems more likely that, despite the similarities between alphavirus and potyvirus leader proteases, they are just mechanistically different. Nonetheless, it cannot be ruled out that P1 proteases can work *in trans* when the right conditions and substrates are presented, opening the door to new functions.

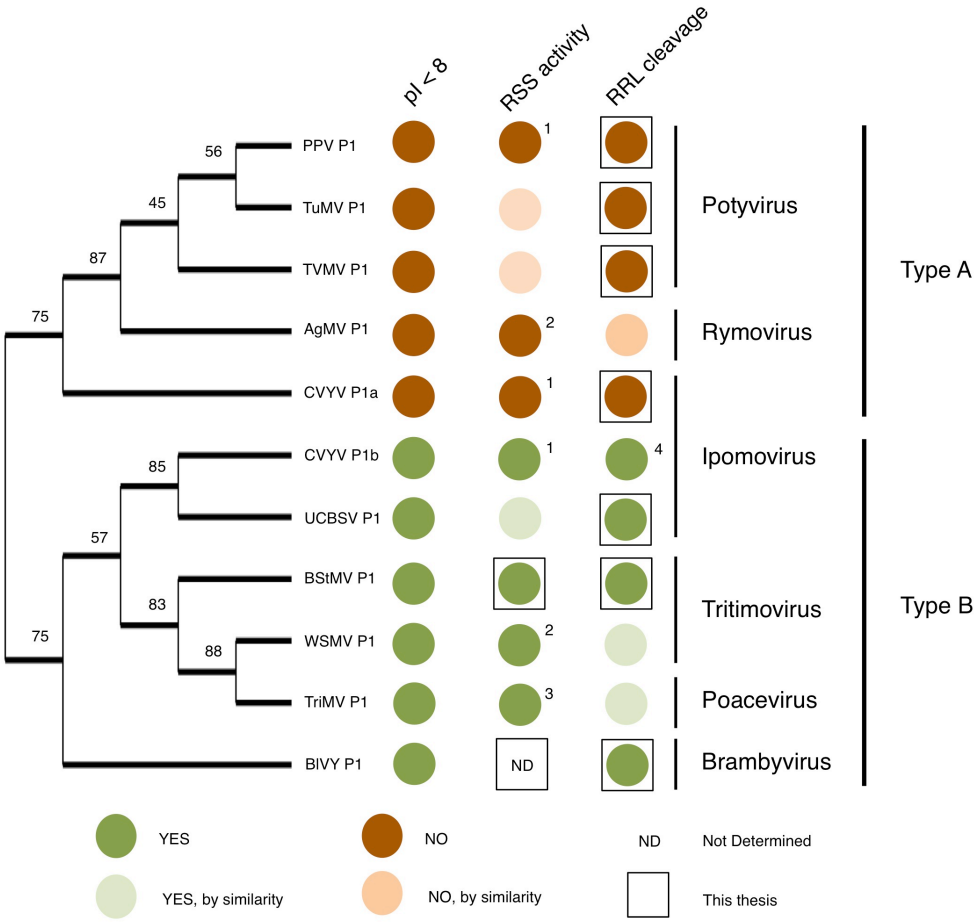

Figure IV.1. Phylogenetic tree and corresponding proteolytic characteristics of Type A and Type B P1 proteins of the *Potyviridae* family. Eleven different sequences representative of the 6 monopartite genera of the family *Potyviridae* that bear P1 proteins were used to build the phylogenetic tree. Bootstrap percentages are shown next to the branches. GenBank information used for the generation of the phylogenetic tree is given in Table II.3. Superscript numbers indicate the following references: [1](Valli et al., 2006), [2](Young et al., 2012), [3](Tatineni et al., 2012), [4](Rodamilans et al., 2013).

IV.2 Type A P1 proteins are involved in host range delimitation

Recent work with a chimeric PPV virus carrying CVYV P1a-P1b (PPV-P1aP1b) instead of P1-HC (Carbonell et al., 2012) revealed a reduced infectivity of this virus in *N. benthamiana* plants compared to wild type PPV or to a chimeric virus carrying P1b sequence instead of P1-HC (PPV-P1b), and we have observed a similar low infectivity in a PPV-P1aHC This suggests a deleterious effect caused by CVYV P1a that might be related to an inability to perform autocleavage in this host, considering that CVYV is not able to infect *Nicotiana* plants.

The inclusion of an NIa cleavage site sequence between a protease-deficient P1 mutant and HC restored infectivity of a TEV clone (Verchot & Carrington, 1995a). Results in this thesis demonstrate that a similar NIa-based strategy significantly enhances infection efficiency of the

PPV-P1aHC chimera in *N. benthamiana* plants (Figure III.5). This effect might be explained by an inactivation of the RNA silencing suppression activity of HC when another protein, such as P1a, remains attached to its N-terminus. Co-expression of a heterologous suppressor is reported to sustain accumulation of viral clones with non-functional suppressor proteins (Garcia-Ruiz et al., 2010). Results of a transient infection in *N. benthamiana* in which infectivity of the PPV chimera that expresses a P1a self-cleavage mutant was rescued by co-delivery of the RSS p19 (Figure III.6) supports the link between P1a self-cleavage defects and silencing suppressor failures. Similar co-expression assays in *N. benthamiana* were performed using transient expression constructs of the corresponding leader proteases. Co-delivery of p19 with each construct permitted detection of fusion products that could not otherwise be observed (Figure III.6). This indicates that the amounts of free HC released by P1a are insufficient to counteract the *N. benthamiana* silencing response, or that non-functional P1a-HC fusion products might have dominant negative effects. Such effects appear to be unlikely, however, since transgenic plants that express the P1-HC cistron can restore viability of TEV clones altered by mutations in the P1 protease (Verchot & Carrington, 1995b). To further support the notion that the lack of P1 processing causes a defect in HC that renders this protein inactive as RNA silencing suppressor, another transient expression experiment was performed in which, similar to the viral construct, an artificial protease cleavage site between P1a and HC was engineered, enabling efficient P1a release from HC and restoring HC RNA silencing suppression activity in *N. benthamiana* (Figure III.7).

Deepening in the relationship between host and P1 processing, experiments with the mentioned viral constructs were performed in *C. sativus*. Taking into account that this is the natural host of CVYV, it was anticipated that PPV chimeric viruses carrying P1a coding sequence would experience a boost in replication, at least at the local inoculation level. Results confirm this prediction showing an opposite outcome to the one obtained in *Nicotiana* hosts (Figure III.8), although none of the tested PPV constructs showed viral accumulation in the upper non-inoculated leaves. In addition, transient RSS assays in cucumber (the CVYV natural host) showed that the P1a-HC construct behaves similarly to the strong suppressor p19, and sustains accumulation of the GFP reporter product (Figure III.9). This observation is consistent with accumulation of the PPV-P1a chimera in this host, which was greater than that of wild-type PPV and somewhat independent of the addition of an NIa cleavage site between P1a and HC. These findings suggest that cucumber provides the co-factor(s) needed for optimal P1a self-cleavage, and directly implicates P1a and other Type A P1 proteases in host range definition, as proposed (Rodamilans et al., 2013;Salvador et al., 2008b;Valli et al., 2007).

The recent discovery that the intrinsically disordered N-terminal region of P1 has an antagonistic effect on the protease activity of the protein and that the C-terminal domain of PPV P1 (P1Pro) is active independent of plant co-factor prompted an experiment to try to define the limits of the interfering domain in order to gain a better understanding on how it might operate. In a previous work, progressive N-terminal deletions allowed to map the border of the minimal cofactor-independent protease domain of PPV P1 at amino acid R165 (Pasin et al., 2014b). An additional series of shorter N-terminal truncations failed to delimit a definite N-terminal antagonistic domain, as N-terminal extensions progressively reduced proteolytic activity in RRL with no detrimental effect in WGE (Figure III.10).

As part of the characterization of the leader proteases inside the *Potyviridae* family and based on the rather high conservation that Type A P1 proteins show on their protease domains (Rodamilans et al., 2013;Valli et al., 2007), experiments were performed intended to find plant cofactor-independent domains similar to PPV P1Pro, in TuMV P1 and CVYV P1a. Results obtained show otherwise (Figure III.11) and implicate the C-terminal part of Type A P1 proteins in host factor mediation. Findings on CVYV P1aPro were confirmed by *in planta* studies (Figure III.16). How PPV P1Pro can overcome this factor requirement is unknown. The results suggest that the protease domain of the potyviral Type A P1 protein demands strict structural features to be active. The still unidentified plant cofactor(s) facilitates this singular folding, which is autonomously achieved by the PPV P1Pro sector. The precision of the PPV P1Pro structure is highlighted by the fact that point mutations, which do not disturb the intrinsic protease activity, makes it dependent of the plant co-factor(s) (Figure III.12).

Experiments that indicate the non-dependence of PPV P1Pro on a host factor have been previously carried out *in vitro* (Pasin, 2015;Pasin et al., 2014b). PPV-P1Pro viral infection in *N. benthamiana* plants displays distinctive features in terms of replication and symptom development and provides evidence that P1Pro release is co-factor-independent. The lack of dependence on a specific co-factor suggests that a P1Pro deletion mutant virus might expand its host range. In agreement with this assumption, *in vivo* transient expression experiments showed efficient P1Pro release from HC in cucumber plants, a non-permissive host of PPV where self-cleavage of the wild type PPV P1 is poor (Figure III.13). Moreover, the mutant virus PPV-P1Pro is able to boost local replication in *C. sativus* leaves indicating that part of the host restriction found by PPV is caused by P1- host co-factor(s) incompatibility (Figure III.14). Replication in the inoculated leaves was confirmed by using a replicase-deficient virus (PPV-P1Pro$_{\Delta GDD}$) and performing RT-qPCR analysis to measure the levels of viral RNA accumulation from both strands (Figure III.15). PPV-P1Pro was also tested in woody hosts. The PPV isolate which

serves as backbone for PPV-P1Pro has been adapted to herbaceous plants and has lost its ability to infect *Prunus* species (Salvador et al., 2008a). This fitness cost might be related to P1 cleavage considering that replacement of the initial part of the virus for the sequence of a wild type PPV-D (PPV-5´BD) rescues infectivity in woody hosts. Experiments performed with three different *Prunus* species indicate that PPV-P1Pro does not improve viral accumulation or movement in these hosts and suggest that limitations of this virus are not related to, or at least are not exclusive of, P1 proteolysis (Figure III.17).

All these results not only highlight the relevance of P1 leader proteinases as host range determinants, but also pose interesting questions about the role of P1 proteins in viral adaptation. As mentioned, P1 is a very moldable proteinase being its N-terminal part the most divergent among proteins in the family, and has suffered numerous recombination events in order to adapt viruses to different hosts (Valli et al., 2007). Could N-terminal deletions be part of this adaptive process? None has been reported in P1 proteins of any member of the *Potyviridae*. However, deletions involved in adaptation to the host, have already been described for the highly variable N-terminal part of PPV CP (Carbonell et al., 2012;Salvador et al., 2008b) and are described for other viruses as well (Sorrell *et al.*, 2009;Tromas *et al.*, 2014;Willcocks *et al.*, 1994). Considering that PPV P1 might acts as a regulator of viral infection (Pasin et al., 2014b), and taking into account that deletions of the N-terminal part of PPV P1 can gradually alter the ability of the protease for self-cleavage, it is reasonable to consider that different deletions on this non-essential gene could modulate viral replication rates and facilitate viral adaptation in specific environments, the same way that PPV-P1Pro is able to overcome hosts incompatibilities and promote local viral amplification in *C. sativus*.

In addition, considering these data, the previous model in which plant factor-P1 protein interaction occurred only through the N-terminal part of the protein has to be revised and adjusted to the new findings (Figure IV.2). Based on the variability of the N-terminal part of Type A P1 proteins (Adams et al., 2005a;Valli et al., 2007;Shukla *et al.*, 1991) it is reasonable to assume that this region is involved in host range definition, but this thesis presents strong evidence that C-terminal part of the protein determine host factor compatibility as well.

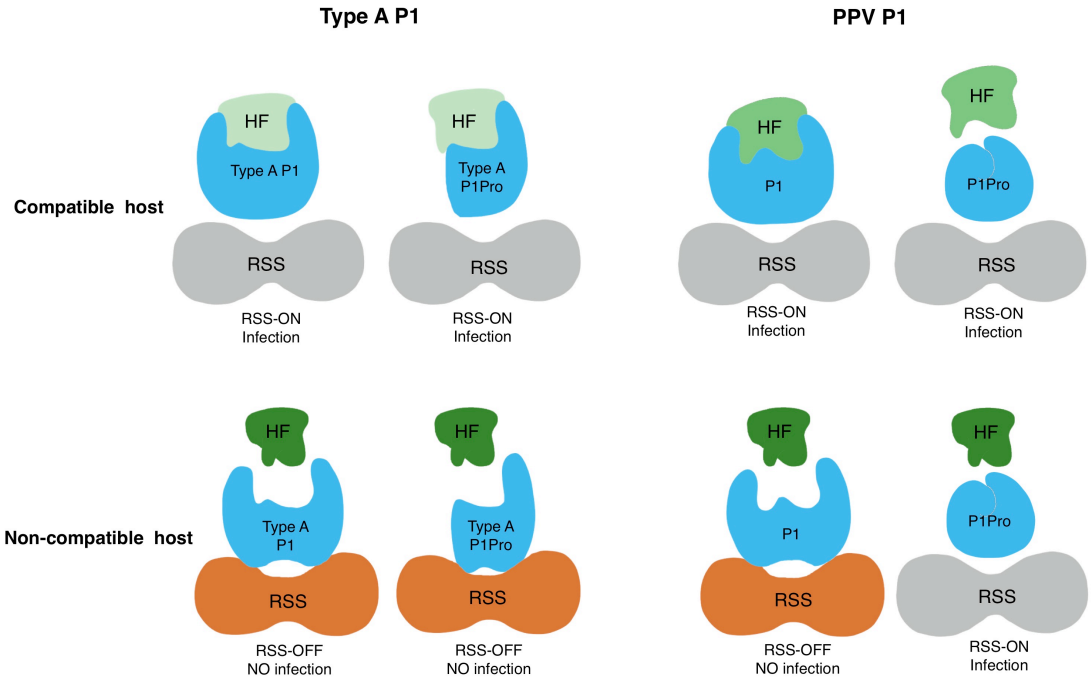

Figure IV.2. Proposed model of Type A P1 proteins and host factor interaction. Usual Type A P1 proteins are depicted on the left. The specific case of Type A PPV P1 is depicted on the right; RSS ON and OFF indicate the active and inactive states of the RNA silencing suppressor; HF is short for host factor and different tones of green indicate different proteins.

IV.3 Searching for the proteolytic plant co-factor(s) of Type A P1 proteins

Proteases in eukaryotic organisms are commonly present in an inactive state to avoid non-specific processing and control their activities in time and space. The best-studied case is the serine protease trypsin, synthesized as the inactive zymogen trypsinogen (Khan & James, 1998). Activation of trypsinogen requires processing of its N-terminal part, a mechanism that might resemble the behavior observed in PPV P1 endopeptidases. In the case of the viral protein, no processing has been reported, but the N-terminal part acts as a negative modulator of the protease activity and the interaction between the host factor and the unfolded N-terminal part of the protein is used as an instrument to regulate various viral functions, such as replication and/or host range delimitation (Pasin et al., 2014b). The mechanism of zymogen activation varies from enzymatic or nonenzymatic cofactor interaction, to a simple change in pH (Khan & James, 1998). In the plant-pathogen interaction, an example of protease activation occurs during infection of *Arabidopsis* by *Pseudomonas syringae*. This bacteria produce the effector protein

AvrRpt2, a cysteine protease that is activated by the host factor chaperone cyclophilin ROC1 as part of a complex innate immune response (Coaker *et al.*, 2005;Coaker *et al.*, 2006). Data on P1 leader proteases also indicate that the proteolytic acitivty of these endopeptidases requires a factor(s) from the plant of protein origin (Verchot et al., 1992).

Identification of the element or elements involved in P1 proteolytic processing is a very challenging task. Several aspects were considered before undertaking purification of the factor(s) from BYL extract. The main concern of this biochemical approach is how the findings of an *in vitro* system translate into live plants. At the moment, no discrepancies have been found between *in vitro* experiments and *in planta* tests in the case of PPV P1, PPV P1Pro, CVYV P1a or CVYV P1aPro. In addition, this methodology presents a crucial advantage: it is a positive approach. This means that the presence of the factor(s) in the starting sample is ensured and that all factors involved in the P1 proteolytic process are subject of identification if multiple elements are involved. This is not the case when the methodology relies on a negative approach such as the more classical use of plant mutant libraries in which the gene(s) of interest have been knocked down. In addition, the likely possibility of gene redundancy can make the efforts of this approach useless.

Before starting the BYL purification process, some experiments were required. The first one consisted on verifying that PPV P1 enzymatic activity could be complemented by WGE when translated in RRL, as it happened in the case of TEV P1 (Verchot et al., 1992) (Figure III.18). Once the result was corroborated, the following steps included production of a BYL extract suitable as PPV P1 proteolytic supplement. As in the previous case, initial trials were positive and not only a viable extract was obtained from tobacco BY-2 cells, but linearity in the P1 proteolytic response could be demonstrated. In addition, the possibility to use non-evacuolated BYL protoplasts for the preparation of the extract increased the initial protein amount obtained from each protoplast preparation and facilitated subsequent factor(s) purification (Figure III.19 and III.20).

A combination of ion exchange and size exclusion chromatography helped to purify an isolated set of proteins that were able to complement PPV P1 proteolytic activity in RRL *in vitro* translation assays (Figures III.21-24). Initially, only two purification steps were included (anion exchange and gel filtration chromatography) and the obtained results gave interesting clues about what the factor(s) might be. Analysis of only specific bands is, however, risky since many proteins present in the supplemented extract are not taken into account. Thus, two additional purification trials were performed including an extra purification step (cation exchange chromatography) that allowed for a better isolation of the plant factor(s). The reduced set of

proteins obtained in the second purification trial (Table III.2) could be cross-referenced with the data obtained from the third purification experiment to get a list of nine candidates likely involved in P1 autocleavage, combined or individually (Table III.3). The high reproducibility of the purification obtaining the plant factor(s) always at the same fraction of the gel filtration column supports the obtained results.

Among the reduced list of proteins identified (Table III.3), three of them stand out as candidates to be involved in P1 cleavage: TUDOR SN-1, HSP70 and HSP89. TUDOR was identified in the initial analysis in bands from fractions showing high P1 processing activity (bands 6 and 7, Table III.1) and was found to be the most abundant protein in fractions 14 of the gel filtration of the other two purification trials. In addition, despite its abundance in these fractions, it was absent from fraction 11 of the third purification trial, which presented no P1 proteolytic activity (Table III.2). On top of that, TUDOR was identified as a P1 interactor in a previous unrelated study (Martínez, 2014 #118), reinforcing its possible involvement with the activity of this protease. HSP70 and HSP89 are both chaperones. These proteins are by definition ideal candidates to facilitate protein folding of the disordered N-terminal P1 region and to favor proteolytic cleavage (Lackie *et al.*, 2017;Hartl & Hayer-Hartl, 2002;Coaker et al., 2005;Coaker et al., 2006). HSP70 is present as TUDOR in the initial band analysis (band 6, Table III.1) as well as in fractions 14 of the gel filtration of subsequent purification trials (Table III.2). In addition, it was also characterized as a possible P1 interactor (Martínez, 2014 #118). However, HSP70 is also present in fractions that show very little or no P1 proteolytic activity, such as band 5 of the first analysis (Table III.1) or fraction 11 of the gel filtration of the third purification trial (Table III.2). This suggests that if HSP70 is a factor helping P1 proteolytic activity, it is not acting alone. On the other hand, HSP89 is present only in fractions 14 of the gel filtration in the second and third purification trials (Table III.2), which fits better with a single activity of this protein as a P1 co-factor.

Final step of the factor(s) identification process is the validation of the candidates. Strategies are based on observing the effect on P1 cleavage of either adding or removing the factor(s) from the experimental system of choice. Initially, the simplest approach was to perform *in vitro* co-translation of P1 and the corresponding candidate and observe the effect of candidate expression on proteolytic activity. Unfortunately, this approach presented a serious setback: the more candidate translated in the system, the less P1HC* product being produced. This, and the fact that the amount of factor(s) needed for P1 complementation is unknown led to the search for a different validation methodology. Subsequent approaches were based on removing factors from BYL supplements in *in vitro* experiments by the use inhibitors (Figure III.25) or downregulating

candidate gene expression in plants by performing VIGS (Figure III.26). None of these tests gave positive results, but this does not rule out any of the tested candidates from being relevant in P1 processing. In the case of the inhibition test, it should be taking into account that blockage of the nuclease function of TUDOR SN-1 protein does not imply that other functions or the protein itself are still valid counterparts for P1 proteinases. In addition, as it was the case of the initial *in vitro* validation test, the amount of factor needed for P1 processing is unknown and therefore, a reduction in nuclease activity might not have a visible effect on P1 proteolytic activity. This same concern can be applied to the VIGS experiment. In this case, candidate genes are not knocked out, but only downregulated and it is possible that the remaining protein expression of the candidates is sufficient for protease complementation. Moreover, variability in downregulation of protein expression among plants complicates the analysis making harder the detection of small variations. Viral infection with TRV, used to silence the candidate genes, is another point of concern since it might interfere unexpectedly with the results. Despite all this, tests are a good starting point for candidate validation and help to elucidate possible setbacks and concerns in the selection of following trials. New validation methods should be explored in order to identify the plant factor(s) involved in P1 processing. These will likely include *in vitro* production of the candidates followed by a single purification step to use them as supplements during RRL translation and over-expression of candidates in *C. sativus* to test their ability to complement P1 processing in a non-permissive host.

The study of P1 proteases is a fascinating topic from several perspectives. Viral proteases are top antiviral targets that have been successfully used in the treatment of diseases such as HCV or AIDS. Control of plant viruses following a similar approach is still behind their animal counterparts, but increased knowledge on plant viral proteases can overcome these limitations. On the other hand, understanding the mechanism of P1 endopeptidases can provide a powerful biotechnological tool opening the door to the engineering of viruses with expanded host ranges or viruses with modulated responses to already susceptible hosts. Moreover, obtaining deep knowledge of P1 proteases and their mechanisms of action is key for the scientific understanding of the *Potyviridae* family. P1 leader proteases seem to be behind the incredible variety and host range success of this viral family. They are the most diverse proteins and their variability is a testimony of their relevance in terms of evolution and adaptation. This thesis considers all these issues and provides information relevant for understanding some of the points mentioned and set up the grounds for advancing in others, broadening our knowledge about P1 proteases, their mechanism of action, the host factors involved in their processing and their possible role in viral evolution.

CONCLUSIONS

V. Conclusions

1. P1 proteases in the *Potyviridae* family can be classified as Type A or Type B based on phylogenetic relationship, isoelectric point, RSS activity and host factor(s) requirement.

2. Type A P1 proteases are host factor(s)-dependent and their processing can limit viral infectivity as shown by the behavior of PPV and the chimeric viruses PPV-P1a and PPV-P1aNIa in *N. benthamiana* and *C. sativus*.

3. PPV P1Pro proteolytic performance *in vitro* can be reproduced in plants enabling processing of the protease in a host-independent manner and promoting viral replication of mutant virus PPV-P1Pro in the non-permissive host, *C. sativus*.

4. Host factor(s) interaction with Type A P1 proteins, such as TuMV P1 and CVYV P1a, involves, not only the N-terminal region of the protein, but also the more conserved C-terminal protease domain.

5. An extract from *N. tabacum* BY-2 cells can be used to complement PPV P1 proteolytic activity during *in vitro* RRL translation.

6. The *N. tabacum* BY-2 cell extract can be fractionated in order to obtain a list of candidates that likely include the factor(s) responsible for P1 proteolytic activity. Cross-referenced of several experiments reduced the initial list to nine candidates pending of validation.

REFERENCES

VI. References

Abraham, A., Menzel, W., Vetten, H. J. and Winter, S. (2012) Analysis of the tomato mild mottle virus genome indicates that it is the most divergent member of the genus *Ipomovirus* (family *Potyviridae*). *Archives of Virology,* **157,** 353-357.

Adams, M. J., Antoniw, J. F. and Beaudoin, F. (2005a) Overview and analysis of the polyprotein cleavage sites in the family *Potyviridae*. *Molecular Plant Pathology,* **6,** 471-487.

Adams, M. J., Antoniw, J. F. and Fauquet, C. M. (2005b) Molecular criteria for genus and species discrimination within the family *Potyviridae*. *Archives of Virology,* **150,** 459-479.

Adams, M. J., Lefkowitz, E. J., King, A. M. Q., Harrach, B., Harrison, R. L., Knowles, N. J.*, et al.* (2017) Changes to taxonomy and the International Code of Virus Classification and Nomenclature ratified by the International Committee on Taxonomy of Viruses (2017). *Archives of Virology,* **162,** 2505-2538.

Agbeci, M., Grangeon, R., Nelson, R. S., Zheng, H. and Laliberte, J. F. (2013) Contribution of host intracellular transport machineries to intercellular movement of turnip mosaic virus. *PLoS Pathogens,* **9,** e1003683.

Aggarwal, M., Dhindwal, S., Kumar, P., Kuhn, R. J. and Tomar, S. (2014) *trans*-Protease activity and structural insights into the active form of the alphavirus capsid protease. *Journal of Virology,* **88,** 12242-12253.

Anandalakshmi, R., Pruss, G. J., Ge, X., Marathe, R., Mallory, A. C., Smith, T. H.*, et al.* (1998) A viral suppressor of gene silencing in plants. *Proceedings of the National Academy of Sciences USA,* **95,** 13079-13084.

Anderson, J., Schiffer, C., Lee, S. and Swanstrom, R. (2009) Viral protease inhibitors. In: *Antiviral Strategies, Handbook of Experimental Pharmacology,* Kräusslich, H. G., Bartenschlager, R., eds, pp 85-110.

Atallah, O. O., Kang, S. H., El-Mohtar, C. A., Shilts, T., Bergua, M. and Folimonova, S. Y. (2016) A 5′-proximal region of the *Citrus tristeza virus* genome encoding two leader proteases is involved in virus superinfection exclusion. *Virology,* **489,** 108-115.

Babe, L. M. and Craik, C. S. (1997) Viral proteases: evolution of diverse structural motifs to optimize function. *Cell,* **91,** 427-430.

Bailey-Elkin, B. A., van Kasteren, P. B., Snijder, E. J., Kikkert, M. and Mark, B. L. (2014) Viral OTU deubiquitinases: a structural and functional comparison. *PLoS Pathogens,* **10,** e1003894.

Bak, A., Cheung, A. L., Yang, C., Whitham, S. A. and Casteel, C. L. (2017) A viral protease relocalizes in the presence of the vector to promote vector performance. *Nature Communications,* **8,** 14493.

Beauchemin, C., Boutet, N. and Laliberté, J. F. (2007) Visualization of the interaction between the precursors of VPg, the viral protein linked to the genome of *Turnip mosaic virus*, and the translation eukaryotic initiation factor iso 4E in planta. *Journal of Virology,* **81,** 775-782.

Calvo, M., Malinowski, T. and García, J. A. (2014) Single amino acid changes in the 6K1-CI region can promote the alternative adaptation of *Prunus*- and *Nicotiana*-propagated *Plum pox virus* C isolates to either host. *Molecular Plant-Microbe Interactions,* **27,** 136-149.

Camborde, L., Planchais, S., Tournier, V., Jakubiec, A., Drugeon, G., Lacassagne, E*., et al.* (2010) The ubiquitin-proteasome system regulates the accumulation of *Turnip yellow mosaic virus* RNA-dependent RNA polymerase during viral infection. *The Plant cell,* **22,** 3142-3152.

Carbonell, A., Dujovny, G., García, J. A. and Valli, A. (2012) The *Cucumber vein yellowing virus* silencing suppressor P1b can functionally replace HCPro in *Plum pox virus* infection in a host-specific manner. *Molecular Plant-Microbe Interactions,* **25,** 151-164.

Carrington, J. C. and Dougherty, W. G. (1987) Small nuclear inclusion protein encoded by a plant potyvirus genome is a protease. *Journal of Virology,* **61,** 2540-2548.

Carrington, J. C., Freed, D. D. and Oh, C. S. (1990) Expression of potyviral polyproteins in transgenic plants reveals three proteolytic activities required for complete processing. *EMBO Journal,* **9,** 1347-1353.

Casteel, C. L., Yang, C., Nanduri, A. C., De Jong, H. N., Whitham, S. A. and Jander, G. (2014) The NIa-Pro protein of *Turnip mosaic virus* improves growth and reproduction of the aphid vector, *Myzus persicae* (green peach aphid). *Plant Journal,* **77,** 653-663.

Cesaratto, F., Burrone, O. R. and Petris, G. (2016) *Tobacco etch virus* protease: a shortcut across biotechnologies. *Journal of Biotechnology,* **231,** 239-249.

Champagne, J., Benhamou, N. and Leclerc, D. (2004) Localization of the N-terminal domain of *Cauliflower mosaic virus* coat protein precursor. *Virology,* **324,** 257-262.

Chase, A. J. and Semler, B. L. (2012) Viral subversion of host functions for picornavirus translation and RNA replication. *Future Virology,* **7,** 179-191.

Chen, K. C., Chiang, C. H., Raja, J. A., Liu, F. L., Tai, C. H. and Yeh, S. D. (2008) A single amino acid of NIapro of *Papaya ringspot virus* determines host specificity for infection of papaya. *Molecular Plant-Microbe Interactions,* **21,** 1046-1057.

Chenon, M., Camborde, L., Cheminant, S. and Jupin, I. (2012) A viral deubiquitylating enzyme targets viral RNA-dependent RNA polymerase and affects viral infectivity. *EMBO Journal,* **31,** 741-753.

Chiu, W., Niwa, Y., Zeng, W., Hirano, T., Kaobayash, H. and Sheen, J. (1996) Engineered GFP as a vital reporter in plant. *Current Biology*, **6**, 325-330.

Clark, V. C., Peter, J. A. and Nelson, D. R. (2013) New therapeutic strategies in HCV: second-generation protease inhibitors. *Liver International,* **33,** 80-84.

Clementz, M. A., Chen, Z., Banach, B. S., Wang, Y., Sun, L., Ratia, K., *et al.* (2010) Deubiquitinating and interferon antagonism activities of coronavirus papain-like proteases. *Journal of Virology,* **84,** 4619-4629.

Coaker, G., Falick, A. and Staskawicz, B. (2005) Activation of a phytopathogenic bacterial effector protein by a eukaryotic cyclophilin. *Science,* **308,** 548-550.

Coaker, G., Zhu, G., Ding, Z., Van Doren, S. R. and Staskawicz, B. (2006) Eukaryotic cyclophilin as a molecular switch for effector activation. *Molecular Microbiology,* **61,** 1485-1496.

Csorba, T., Kontra, L. and Burgyan, J. (2015) viral silencing suppressors: Tools forged to fine-tune host-pathogen coexistence. *Virology,* **479-480,** 85-103.

Cui, H. and Wang, A. (2016) *Plum pox virus* 6K1 protein is required for viral replication and targets the viral replication complex at the early stage of infection. *Journal of Virology,* **90,** 5119-5131.

Mejia, M. V., Hiebert, E., Purcifull, D. E., Thornbury, D. W. and Pirone, T. P. (1985) Identification of potyviral amorphous inclusion protein as a nonstructural, virus-specific protein related to helper component. *Virology,* **142,** 34-43.

Dereeper, A., Audic, S., Claverie, J. M. and Blanc, G. (2010) BLAST-EXPLORER helps you building datasets for phylogenetic analysis. *BMC Evolutionary Biology,* **10,** 8.

Desbiez, C. and Lecoq, H. (2004) The nucleotide sequence of *Watermelon mosaic virus* (WMV, *Potyvirus*) reveals interspecific recombination between two related potyviruses in the 5' part of the genome. *Archives of Virology,* **149,** 1619-1632.

Dolja, V. V. (1994) Molecular biology and evolution of *Closteroviruses*: Sophisticated build-up of large RNA genomes. *Annual Review of Phytopathology,* **32,** 261-285.

Dolja, V. V. and Koonin, E. V. (2011) Common origins and host-dependent diversity of plant and animal viromes. *Current Opinion in Virology,* **1,** 322-331.

Eickbush, T. H. and Jamburuthugoda, V. K. (2008) The diversity of retrotransposons and the properties of their reverse transcriptases. *Virus Research,* **134,** 221-234.

Fernández-Fernández, M. R., Mouriño, M., Rivera, J., Rodríguez, F., Plana-Durán, J. and García, J. A. (2001) Protection of rabbits against *Rabbit hemorrhagic disease virus* by immunization with the VP60 protein expressed in plants with a potyvirus-based vector. *Virology,* **280,** 283-291.

Fernandez-Rodriguez, J. and Voigt, C. A. (2016) Post-translational control of genetic circuits using *Potyvirus* proteases. *Nucleic Acids Research,* **44,** 6493-6502.

Firth, A. E. and Brierley, I. (2012) Non-canonical translation in RNA viruses. *Journal of General Virology,* **93,** 1385-1409.

Foster, G. D. and Mills, P. R. (1992) Translation of *Potato virus S* RNA *in vitro*: evidence of protein processing. *Virus Genes,* **6,** 47-52.

Gable, J. E., Lee, G. M., Jaishankar, P., Hearn, B. R., Waddling, C. A., Renslo, A. R.*, et al.* (2014) Broad-spectrum allosteric inhibition of herpesvirus proteases. *Biochemistry,* **53,** 4648-4660.

Gallo, A. (2017) Conexiones entre la replicación del RNA, su empaquetamiento y la proteína HC en un potyvirus. PhD thesis. Universidad Autónoma de Madrid.

García, J. A., Cervera, M. T., Riechmann, J. L. and López-Otín, C. (1993) Inhibitory effects of human cystatin C on *Plum pox potyvirus* proteases. *Plant Molecular Biology,* **22,** 697-701.

García, J. A., Glasa, M., Cambra, M. and Candresse, T. (2014) *Plum pox virus* and sharka: a model potyvirus and a major disease. *Molecular Plant Pathology,* **15,** 226-241.

Garcia-Ruiz, H., Takeda, A., Chapman, E. J., Sullivan, C. M., Fahlgren, N., Brempelis, K. J.*, et al.* (2010) Arabidopsis RNA-dependent RNA polymerases and dicer-like proteins in antiviral defense and small interfering RNA biogenesis during *Turnip mosaic virus* infection. *Plant Cell,* **22,** 481-496.

Gholizadeh, A., Santha, I. M., Kohnehrouz, B. B., Lodha, M. L. and Kapoor, H. C. (2005) Cystatins may confer viral resistance in plants by inhibition of a virus-induced cell death phenomenon in which cysteine proteinases are active: cloning and molecular characterization of a cDNA encoding cysteine-proteinase inhibitor (celostatin) from *Celosia cristata* (crested cock's comb). *Biotechnology and Applied Biochemistry,* **42,** 197-204.

Gibson, D., Young, L., Chuang, R., Venter, J., Hytchison III, C. and Smih, H. (2009) Enzymatic assembly of DNA molecules up to several hundred kilobases. *Nature Methods*, **6**, 343-345.

Giner, A., Lakatos, L., García-Chapa, M., López-Moya, J. J. and Burgyan, J. (2010) Viral protein inhibits RISC activity by argonaute binding through conserved WG/GW motifs. *PLoS Pathogens,* **6,** e1000996.

Goldbach, R., Le Gall, O. and Wellink, J. (1991) Alpha-like viruses in plants. *Seminars in Virology,* **2,** 19-25.

Govind, K., Mäkinen, K. and Savithri, H. S. (2012) *Sesbania mosaic virus* (SeMV) infectious clone: possible mechanism of 3' and 5' end repair and role of polyprotein processing in viral replication. *PLoS One,* **7,** e31190.

Guarné, A., Tormo, J., Kirchweger, R., Pfistermueller, D., Fita, I. and Skern, T. (1998) Structure of the foot-and-mouth disease virus leader protease: a papain-like fold adapted for self-processing and eIF4G recognition. *EMBO Journal,* **17,** 7469-7479.

Guo, B., Lin, J. and Ye, K. (2011) Structure of the autocatalytic cysteine protease domain of potyvirus helper-component proteinase. *Journal of Biological Chemistry,* **286,** 21937-21943.

Guo, H. S., López-Moya, J. J. and García, J. A. (1999) Mitotic stability of infection-induced resistance to plum pox potyvirus associated with transgene silencing and DNA methylation. *Molecular Plant-Microbe Interactions,* **12,** 103-111.

Gutiérrez-Campos, R., Torres-Acosta, J. A., Pérez-Martínez, J. J. and Gómez-Lim, M. A. (2001) Pleiotropic effects in transgenic tobacco plants expressing the oryzacystatin I gene. *HortScience,* **36,** 118-119.

Gutierrez-Campos, R., Torres-Acosta, J. A., Saucedo-Arias, L. J. and Gomez-Lim, M. A. (1999) The use of cysteine proteinase inhibitors to engineer resistance against potyviruses in transgenic tobacco plants. *Nature Biotechnology,* **17,** 1223-1226.

Habib, H. a. F., M. (2007) Plant protease inhibitors: a defense strategy in plants. *Biotechnology and Molecular Biology Review,* **2,** 68-85.

Hagiwara-Komoda, Y., Choi, S. H., Sato, M., Atsumi, G., Abe, J., Fukuda, J.*, et al.* (2016) Truncated yet functional viral protein produced via RNA polymerase slippage implies underestimated coding capacity of RNA viruses. *Scientific Reports,* **6,** 21411.

Halgren, A., Tzanetakis, I. E. and Martin, R. R. (2007) Identification, characterization, and detection of *Black raspberry necrosis virus*. *Phytopathology,* **97,** 44-50.

Han, H. E., Sellamuthu, S., Shin, B. H., Lee, Y. J., Song, S., Seo, J. S., *et al.* (2010) The nuclear inclusion a (NIa) protease of *Turnip mosaic virus* (TuMV) cleaves amyloid-ß. *PLoS One,* **5,** e15645.

Hara-Nishimura, I. and Hatsugai, N. (2011) The role of vacuole in plant cell death. *Cell Death Differentiation,* **18,** 1298-1304.

Hartl, F. U. and Hayer-Hartl, M. (2002) Molecular chaperones in the cytosol: from nascent chain to folded protein. *Science,* **295,** 1852-1858.

Haseloff, J., Siemering, K. R., Prasher, D. C., and Hodge, S. (1997) Removal of a cryptic intron and subcellular localization of green fluorescent protein are required to mark transgenic Arabidopsis plants brightly. *Proceeding of National Academy of Sciences USA*, 94, 2122-2127.

Hehn, A., Fritsch, C., Richards, K. E., Guilley, H. and Jonard, G. (1997) Evidence for *in vitro* and *in vivo* autocatalytic processing of the primary translation product of *Beet necrotic yellow vein virus* RNA 1 by a papain-like proteinase. *Archives of Virology,* **142,** 1051-1058.

Heinlein, M. (2015) Plant virus replication and movement. *Virology,* **479-480,** 657-671.

Ingr, M., Uhlikova, T., Strisovsky, K., Majerova, E. and Konvalinka, J. (2003) Kinetics of the dimerization of retroviral proteases: the "fireman's grip" and dimerization. *Protein Science,* **12,** 2173-2182.

Ivanov, K. I., Eskelin, K., Lõhmus, A. and Mäkinen, K. (2014) Molecular and cellular mechanisms underlying potyvirus infection. *Journal of general virology,* **95,** 1415-1429.

Jakubiec, A., Drugeon, G., Camborde, L. and Jupin, I. (2007) Proteolytic processing of *Turnip yellow mosaic virus* replication proteins and functional impact on infectivity. *Journal of Virology,* **81,** 11402-11412.

Jakubiec, A., Notaise, J., Tournier, V., Hericourt, F., Block, M. A., Drugeon, G., *et al.* (2004) Assembly of *Turnip yellow mosaic virus* replication complexes: interaction between the proteinase and polymerase domains of the replication proteins. *Journal of Virology,* **78,** 7945-7957.

Janssen, D., Martín, G., Velasco, L., Gómez, P., Segundo, E., Ruiz, L., *et al.* (2005) Absence of a coding region for the helper component-proteinase in the genome of cucumber vein yellowing virus, a whitefly-transmitted member of the *Potyviridae*. *Archives of Virology,* **150,** 1439-1447.

Jupin, I., Ayach, M., Jomat, L., Fieulaine, S. and Bressanelli, S. (2017) A mobile loop near the active site acts as a switch between the dual activities of a viral protease/deubiquitinase. *PLoS Pathogens,* **13,** e1006714.

Karpe, Y. A. and Lole, K. S. (2011) Deubiquitination activity associated with hepatitis E virus putative papain-like cysteine protease. *Journal of General Virology,* **92,** 2088-2092.

Karsies, A., Merkle, T., Szurek, B., Bonas, U., Hohn, T. and Leclerc, D. (2002) Regulated nuclear targeting of *Cauliflower mosaic virus*. *Journal of General Virology,* **83,** 1783-1790.

Kashiwazaki, S., Minobe, Y. and Hibino, H. (1991) Nucleotide sequence of *Barley yellow mosaic virus* RNA 2. *Journal of General Virology,* **72** 995-999.

Kashiwazaki, S., Minobe, Y., Omura, T. and Hibino, H. (1990) Nucleotide sequence of *Barley yellow mosaic virus* RNA 1: a close evolutionary relationship with potyviruses. *Journal of General Virology,* **72,** 2781-2790.

Kasschau, K. D. and Carrington, J. C. (1998) A counterdefensive strategy of plant viruses: suppression of posttranscriptional gene silencing. *Cell,* **95,** 461-470.

Khan, A. R. and James, M. N. (1998) Molecular mechanisms for the conversion of zymogens to active proteolytic enzymes. *Protein Science,* **7,** 815-836.

Kim, S. H., Qi, D., Ashfield, T., Helm, M. and Innes, R. W. (2016) Using decoys to expand the recognition specificity of a plant disease resistance protein. *Science,* **351,** 684-687.

Kim, T. K., Han, H. E., Kim, H., Lee, J. E., Choi, D., Park, W. J.*, et al.* (2012) Expression of the plant viral protease NIa in the brain of a mouse model of Alzheimer's disease mitigates Aß pathology and improves cognitive function. *Experimental and Molecular Medicine,* **44,** 740-748.

Kondo, T. and Fujita, T. (2012) Complete nucleotide sequence and construction of an infectious clone of *Chinese yam necrotic mosaic virus* suggest that macluraviruses have the smallest genome among members of the family *Potyviridae*. *Archives of Virology,* **157,** 2299-2307.

Konvalinka, J., Krausslich, H. G. and Muller, B. (2015) Retroviral proteases and their roles in virion maturation. *Virology,* **479-480,** 403-417.

Koonin, E. V. (1991) The phylogeny of RNA-dependent RNA polymerases of positive-strand RNA viruses. *Journal of General Virology,* **72,** 2197-2206.

Krupovic, M. and Koonin, E. V. (2017) Multiple origins of viral capsid proteins from cellular ancestors. *Proceedings of the National Academy of Sciences USA,* **114,** 2401-2410.

Lackie, R. E., Maciejewski, A., Ostapchenko, V. G., Marques-Lopes, J., Choy, W. Y., Duennwald, M. L., *et al.* (2017) The Hsp70/Hsp90 chaperone machinery in neurodegenerative diseases. *Frontiers in Neuroscience,* **11,** 254.

Lackner, T., Muller, A., Pankraz, A., Becher, P., Thiel, H. J., Gorbalenya, A. E., *et al.* (2004) Temporal modulation of an autoprotease is crucial for replication and pathogenicity of an RNA virus. *Journal of Virology,* **78,** 10765-10775.

Lackner, T., Thiel, H. J. and Tautz, N. (2006) Dissection of a viral autoprotease elucidates a function of a cellular chaperone in proteolysis. *Proceedings of the National Academy of Sciences USA,* **103,** 1510-1515.

Laco, G. S., Kent, S. B. and Beachy, R. N. (1995) Analysis of the proteolytic processing and activation of the *Rice tungro bacilliform virus* reverse transcriptase. *Virology,* **208,** 207-214.

Lawrence, D. M., Rozanov, M. N. and Hillman, B. I. (1995) Autocatalytic processing of the 223-kDa protein of *Blueberry scorch carlavirus* by a papain-like proteinase. *Virology,* **207,** 127-135.

Li, J., Feng, Z., Wu, J., Huang, Y., Lu, G., Zhu, M., *et al.* (2015) Structure and function analysis of nucleocapsid protein of tomato spotted wilt virus interacting with RNA using homology modeling. *Journal of Biological Chemistry,* **290,** 3950-3961.

Li, X., Halpin, C. and Ryan, M. D. (2007) A novel cleavage site within the *Potato leafroll virus* P1 polyprotein. *Journal of General Virology,* **88,** 1620-1623.

Li, X., Ryan, M. D. and Lamb, J. W. (2000) *Potato leafroll virus* protein P1 contains a serine proteinase domain. *Journal of General Virology,* **81,** 1857-1864.

Liu, Y., Schiff, M. and Dinesh-Kumar, S. P. (2002a) Virus-induced gene silencing in tomato. *Plant Journal,* **31,** 777-786.

Liu, Y., Schiff, M., Marathe, R. and Dinesh-Kumar, S. P. (2002b) Tobacco Rar1, EDS1 and NPR1/NIM1 like genes are required for N-mediated resistance to *Tobacco mosaic virus*. *Plant Journal,* **30,** 415-429.

Liu, Y., Zhu, Z., Zhang, M. and Zheng, H. (2015) Multifunctional roles of leader protein of *Foot-and-mouth disease viruses* in suppressing host antiviral responses. *Veterinary Research,* **46,** 127.

Liu, Y. P., Peremyslov, V. V., Medina, V. and Dolja, V. V. (2009) Tandem leader proteases of *Grapevine leafroll-associated virus-2*: host-specific functions in the infection cycle. *Virology,* **383,** 291-299.

Lombardi, C., Ayach, M., Beaurepaire, L., Chenon, M., Andreani, J., Guerois, R.*, et al.* (2013) A compact viral processing proteinase/ubiquitin hydrolase from the OTU family. *PLoS Pathogens,* **9,** e1003560.

Lucini, C. (2004) Expresíon de proteínas heterólogas en plantas por medio del virus de la sharka (PPV). PhD thesis. Universidad Politécnica de Madrid.

Maliogka, V. I., Salvador, B., Carbonell, A., Sáenz, P., San León, D., Oliveros, J. C.*, et al.* (2012) Virus variants with differences in the P1 protein coexist in a *Plum pox virus* population and display particular host-dependent pathogenicity features. *Molecular Plant Pathology,* **13,** 877-886.

Mann, K. S., Walker, M. and Sanfaçon, H. (2017) Identification of cleavage sites recognized by the 3C-like cysteine protease within the two polyproteins of *Strawberry mottle virus*. *Frontiers in Microbiology,* **8,** 745.

Marmey, P., Rojas-Mendoza, A., de Kochko, A., Beachy, R. N. and Fauquet, C. M. (2005) Characterization of the protease domain of *Rice tungro bacilliform virus* responsible for the processing of the capsid protein from the polyprotein. *Virology Journal,* **2,** 33.

Martínez, F. and Daròs, J. A. (2014) *Tobacco etch virus* protein P1 traffics to the nucleolus and associates with the host 60S ribosomal subunits during infection. *Journal of Virology,* **88,** 10725-10737.

Martínez, F., Rodrigo, G., Aragonés, V., Ruiz, M., Lodewijk, I., Fernández, U.*, et al.* (2016) Interaction network of *Tobacco etch potyvirus* NIa protein with the host proteome during infection. *BMC Genomics,* **17,** 87.

Mbanzibwa, D. R., Tian, Y., Mukasa, S. B. and Valkonen, J. P. (2009) *Cassava brown streak virus* (*Potyviridae*) encodes a putative Maf/HAM1 pyrophosphatase implicated in reduction of mutations and a P1 proteinase that suppresses RNA silencing but contains no HC-Pro. *Journal of Virology,* **83,** 6934-6940.

McGavin, W. J., McMenemy, L. S. and MacFarlane, S. A. (2010) The complete sequence of a UK strain of *Black raspberry necrosis virus*. *Viology*, **155**, 1987-1899.

Merits, A., Rajamäki, M. L., Lindholm, P., Runeberg-Roos, P., Kekarainen, T., Puustinen, P.*, et al.* (2002) Proteolytic processing of potyviral proteins and polyprotein processing intermediates in insect and plant cells. *Journal of General Virology,* **83,** 1211-1221.

Mestre, P., Brigneti, G. and Baulcombe, D. C. (2000) An *Ry*-mediated resistance response in potato requires the intact active site of the NIa proteinase from *Potato virus Y*. *Plant Journal,* **23,** 653-661.

Mestre, P., Brigneti, G., Durrant, M. C. and Baulcombe, D. C. (2003) *Potato virus Y* NIa protease activity is not sufficient for elicitation of *Ry*-mediated disease resistance in potato. *Plant Journal,* **36,** 755-761.

Mingot, A., Valli, A., Rodamilans, B., San León, D., Baulcombe, D. C., García, J. A.*, et al.* (2016) The P1N-PISPO trans-frame gene of *Sweet potato feathery mottle potyvirus* is produced during virus infection and functions as an RNA silencing suppressor. *Journal of Virology,* **90,** 3543-3557.

Miras, M., Miller, W. A., Truniger, V. and Aranda, M. A. (2017) Non-canonical translation in plant RNA viruses. *Frontiers in Plant Science,* **8,** 494.

Moelling, K. (2012) Are viruses our oldest ancestors? *EMBO Reports,* **13,** 1033.

Moreno, M., Brandwagt, B. F., Shaw, J. G. and Rodríguez-Cerezo, E. (1999) Infectious virus in transgenic plants inoculated with a nonviable, P1-proteinase defective mutant of a potyvirus. *Virology,* **257,** 322-329.

Murota, K., Hagiwara-Komoda, Y., Komoda, K., Onouchi, H., Ishikawa, M. and Naito, S. (2011) Arabidopsis cell-free extract, ACE, a new in vitro translation system derived from *Arabidopsis* callus cultures. *Plant Cell Physiology,* **52,** 1443-1453.

Nair, S. and Savithri, H. S. (2010a) Natively unfolded nucleic acid binding P8 domain of SeMV polyprotein 2a affects the novel ATPase activity of the preceding P10 domain. *FEBS Letters,* **584,** 571-576.

Nair, S. and Savithri, H. S. (2010b) Processing of SeMV polyproteins revisited. *Virology,* **396,** 106-117.

Olspert, A., Chung, B. Y., Atkins, J. F., Carr, J. P. and Firth, A. E. (2015) Transcriptional slippage in the positive-sense RNA virus family *Potyviridae*. *EMBO Reports,* **16,** 995-1004.

Pakdel, A., Mounier, C., Klein, E., Hleibieh, K., Monsion, B., Mutterer, J.*, et al.* (2015) On the interaction and localization of the *Beet necrotic yellow vein virus* replicase. *Virus Research,* **196,** 94-104.

Pan, Q. Y. (2016) Molecular characterization of the potyviral first protein (P1 protein). PhD thesis. University of Western Ontario.

Pasin, F. (2015) The *Potyviridae* P1 protease modulates viral replication and host defense responses. PhD thesis. Universidad Autónoma de Madrid.

Pasin, F., Kulasekaran, S., Natale, P., Simón-Mateo, C. and García, J. A. (2014a) Rapid fluorescent reporter quantification by leaf disc analysis and its application in plant-virus studies. *Plant methods,* **10,** 22.

Pasin, F., Simón-Mateo, C. and García, J. A. (2014b) The hypervariable amino-terminus of p1 protease modulates potyviral replication and host defense responses. *PLoS Pathogens,* **10,** e1003985.

Peng, C. W., Napuli, A. J. and Dolja, V. V. (2003) Leader proteinase of *Beet yellows virus* functions in long-distance transport. *Journal of Virology,* **77,** 2843-2849.

Peng, C. W., Peremyslov, V. V., Mushegian, A. R., Dawson, W. O. and Dolja, V. V. (2001) Functional specialization and evolution of leader proteinases in the family *Closteroviridae*. *Journal of Virology,* **75,** 12153-12160.

Peterson-Burch, B. D. and Voytas, D. F. (2002) Genes of the *Pseudoviridae* (Ty1/copia retrotransposons). *Molecular Biology and Evolution,* **19,** 1832-1845.

Prüfer, D., Kawchuk, L., Monecke, M., Nowok, S., Fischer, R. and Rohde, W. (1999) Immunological analysis of *Potato leafroll luteovirus* (PLRV) P1 expression identifies a 25 kDa RNA-binding protein derived via P1 processing. *Nucleic Acids Research,* **27,** 421-425.

Racaniello, V. R. (2001) *Picornaviridae:* the viruses and their replication. Philadelphia, USA: Lippincott Williams & Wilkins.

Rausalu, K., Utt, A., Quirin, T., Varghese, F. S., Zusinaite, E., Das, P. K.*, et al.* (2016) Chikungunya virus infectivity, RNA replication and non-structural polyprotein processing depend on the nsP2 protease´s active site cysteine residue. *Scientific Reports,* **6,** 37124.

Raut, R., Beesetti, H., Tyagi, P., Khanna, I., Jain, S. K., Jeankumar, V. U.*, et al.* (2015) A small molecule inhibitor of dengue virus type 2 protease inhibits the replication of all four dengue virus serotypes in cell culture. *Virology Journal,* **12,** 16.

Rawlings, N. D., Waller, M., Barrett, A. J. and Bateman, A. (2014) MEROPS: the database of proteolytic enzymes, their substrates and inhibitors. *Nucleic Acids Research,* **42,** 503-509.

Restrepo-Hartwig, M. A. and Carrington, J. C. (1994) The *Tobacco etch potyvirus* 6-kilodalton protein is membrane associated and involved in viral replication. *Journal of Virology,* **68,** 2388-2397.

Revers, F. and García, J. A. (2015) Molecular biology of potyviruses. *Advances in Virus Research,* **92,** 101-199.

Riechmann, J. L., Cervera, M. T. and Garcia, J. A. (1995) Processing of the *Plum pox virus* polyprotein at the P3-6K1 junction is not required for virus viability. *Journal of General Virology,* **76** 951-956.

Rodamilans, B., Valli, A. and García, J. A. (2013) Mechanistic divergence between P1 proteases of the family *Potyviridae*. *Journal of General Virology,* **94,** 1407-1414.

Rodamilans, B., Valli, A., Mingot, A., San León, D., Baulcombe, D., López-Moya, J. J.*, et al.* (2015) RNA polymerase slippage as a mechanism for the production of frameshift gene products in plant viruses of the *Potyviridae* family. *Journal of Virology,* **89,** 6965-6967.

Rohila, J., Chen, M., Cerny, R. and Fromm, M. (2004a) Improved tandem affinity purification tag and methods forisolation of protein heterocomplexes from plants. *Plant Journal*, **38**, 172-181.

Rohozkova, J. and Navratil, M. (2011) P1 peptidase – a mysterious protein of family *Potyviridae*. *Journal of Biosciences,* **36,** 189-200.

Roossinck, M. J., Sabanadzovic, S., Okada, R. and Valverde, R. A. (2011) The remarkable evolutionary history of endornaviruses. *Journal of General Virology,* **92,** 2674-2678.

Rozanov, M. N., Drugeon, G. and Haenni, A. L. (1995) Papain-like proteinase of *Turnip yellow mosaic virus*: a prototype of a new viral proteinase group. *Archives of Virology,* **140,** 273-288.

Sabanadzovic, S., Wintermantel, W. M., Valverde, R. A., McCreight, J. D. and Aboughanem-Sabanadzovic, N. (2016) *Cucumis melo endornavirus*: Genome organization, host range and co-divergence with the host. *Virus Research,* **214,** 49-58.

Salvador, B., Delgadillo, M. O., Sáenz, P., García, J. A. and Simón-Mateo, C. (2008a) Identification of *Plum pox virus* pathogenicity determinants in herbaceous and woody hosts. *Molecular Plant-Microbe Interactions,* **21,** 20-29.

Salvador, B., Sáenz, P., Yángüez, E., Quiot, J. B., Quiot, L., Delgadillo, M. O.*, et al.* (2008b) Host-specific effect of P1 exchange between two potyviruses. *Molecular Plant Pathology,* **9,** 147-155.

Sánchez, F., Martínez-Herrera, D., Aguilar, I. and Ponz, F. (1998) Infectivity of *Turnip mosaic potyvirus* cDNA clones and transcripts on the systemic host *Arabidopsis thaliana* and local lesion hosts. *Virus Research,* **55,** 207-219.

Satheshkumar, P. S., Lokesh, G. L. and Savithri, H. S. (2004) Polyprotein processing: *cis* and *trans* proteolytic activities of *Sesbania mosaic virus* serine protease. *Virology,* **318,** 429-438.

Sawicki, D. L. and Sawicki, S. G. (1994) *Alphavirus* positive and negative strand RNA synthesis and the role of polyproteins in formation of viral replication complexes. *Archives of Virology* **9,** 393-405.

Schaad, M. C., Jensen, P. E. and Carrington, J. C. (1997) Formation of plant RNA virus replication complexes on membranes: role of an endoplasmic reticulum-targeted viral protein. *EMBO Journal,* **16,** 4049-4059.

Schneider, C., Rasband, W. and Eliceiri, K. (2012) NIH Image to ImageJ: 25 years of image analysis. *Nature Methods*, **9**, 671-675.

Seipelt, J., Guarné, A., Bergmann, E., James, M., Sommergruber, W., Fita, I.*, et al.* (1999) The structures of picornaviral proteinases. *Virus Research,* **62,** 159-168.

Senthil-Kumar, M. and Mysore, K. S. (2014) *Tobacco rattle virus*-based virus-induced gene silencing in *Nicotiana benthamiana*. *Nature Protocols,* **9,** 1549-1562.

Shamsi, T. N., Parveen, R. and Fatima, S. (2016) Characterization, biomedical and agricultural applications of protease inhibitors: A review. *International Journal of Biological Macromolecules,* **91,** 1120-1133.

Shan, H., Pasín, F., Simón-Mateo, C., García, J. A. and Rodamilans, B. (2018) Truncation of a P1 leader proteinase facilitates potyvirus replication in a non-permissive host. Submitted to publication. *Molecular Plant Pathology,* **in press**.

Shan, H., Pasín, F., Valli, A., Castillo, C., Rajulu, C., Carbonell, A.*, et al.* (2015) The *Potyviridae* P1a leader protease contributes to host range specificity. *Virology,* **476,** 264-270.

Shevchenko, A., Wilm, M., Vorm, O. and Mann, M. (1996) Mass spectrometric sequencing of proteins silver-stained polyacrylamide gels. *Analytical chemistry,* **68,** 850-858.

Shi, X., Botting, C. H., Li, P., Niglas, M., Brennan, B., Shirran, S. L.*, et al.* (2016) *Bunyamwera orthobunyavirus* glycoprotein precursor is processed by cellular signal peptidase and signal peptide peptidase. *Proceedings of the National Academy of Sciences USA,* **113,** 8825-8830.

Shi, Y., Chen, J., Hong, X., Chen, J. and Adams, M. J. (2007) A potyvirus P1 protein interacts with the Rieske Fe/S protein of its host. *Moleculor Plant Pathology,* **8,** 785-790.

Shukla, D. D., Frcnkel, M. J. and Ward, C. W. (1991) Structure and function of the potyvirus genome with special reference to the coat protein coding region. *Canada Journal of Plant Pathology*, **13**, 178-191.

Sihelská, N., Glasa, M. and Šubr, Z. W. (2017) Host preference of the major strains of *Plum pox virus* –Opinions based on regional and world-wide sequence data. *Journal of Integrative Agriculture*, **16**, 510-515.

Simmonds, P., Adams, M. J., Benko, M., Breitbart, M., Brister, J. R., Carstens, E. B., *et al.* (2017) Consensus statement: virus taxonomy in the age of metagenomics. *Nature Reviews Microbiology,* **15,** 161-168.

Sõmera, M., Sarmiento, C. and Truve, E. (2015) Overview on sobemoviruses and a proposal for the creation of the family *Sobemoviridae*. *Viruses,* **7,** 3076-3115.

Sorrell, E. M., Wan, H., Araya, Y., Song, H. and Perez, D. R. (2009) Minimal molecular constraints for respiratory droplet transmission of an *Avian-human H9N2 influenza A virus*. *Proceedings of the National Academy of Sciences UAS,* **106,** 7565-7570.

Spall, V. E., Shanks, M. and Lomonossoff, G. P. (1997) Polyprotein processing as a strategy for gene expression in RNA viruses. *Seminars in Virology,* **8,** 15-23.

Stewart, L. R., Jarugula, S., Zhao, Y., Qu, F. and Marty, D. (2017) Identification of a *Maize chlorotic dwarf virus* silencing suppressor protein. *Virology,* **504,** 88-95.

Sun, D., Chen, S., Cheng, A. and Wang, M. (2016) Roles of the picornaviral 3C proteinase in the viral life cycle and host cells. *Viruses,* **8,** 82.

Susaimuthu, J., Tzanetakis, I. E., Gergerich, R. C. and Martin, R. R. (2008) A member of a new genus in the *Potyviridae* infects rubus. *Virus Research,* **131,** 145-151.

Tatineni, S. and French, R. (2016) The coat protein and NIa protease of two *Potyviridae* family members independently confer superinfection exclusion. *Journal of Virology,* **90,** 10886-10905.

Tatineni, S., McMechan, A. J., Hein, G. L. and French, R. (2010) Efficient and stable expression of GFP through *Wheat streak mosaic virus*-based vectors in cereal hosts using a range of cleavage sites: formation of dense fluorescent aggregates for sensitive virus tracking. *Virology,* **410,** 268-281.

Tatineni, S., Qu, F., Li, R., Morris, T. J. and French, R. (2012) *Triticum mosaic poacevirus* enlists P1 rather than HC-Pro to suppress RNA silencing-mediated host defense. *Virology,* **433,** 104-115.

Thole, V. and Hull, R. (1998) *Rice tungro spherical virus* polyprotein processing: identification of a virus-encoded protease and mutational analysis of putative cleavage sites. *Virology,* **247,** 106-114.

Thompson, J. R., Kamath, N. and Perry, K. L. (2014) An evolutionary analysis of the *Secoviridae* family of viruses. *PLoS One,* **9,** e106305.

Tong, L. (2002) Viral proteases. *Chemical Reviews,* **102,** 4609-4626.

Torruella, M., Gordon, K. and Hohn, T. (1989) *Cauliflower mosaic virus* produces an aspartic proteinase to cleave its polyproteins. *EMBO Journal,* **8,** 2819-2825.

Tran, P. T., Widyasari, K., Park, J. Y. and Kim, K. H. (2017) Engineering an auto-activated R protein that is in vivo activated by a viral protease. *Virology,* **510,** 242-247.

Tromas, N., Zwart, M. P., Lafforgue, G. and Elena, S. F. (2014) Within-host spatiotemporal dynamics of plant virus infection at the cellular level. *PLoS Genetics,* **10,** e1004186.

Untiveros, M., Olspert, A., Artola, K., Firth, A. E., Kreuze, J. F. and Valkonen, J. P. (2016) A novel *Sweet potato potyvirus* ORF is expressed via polymerase slippage and suppresses RNA silencing. *Molecular Plant Pathology,* **17,** 1111-1123.

Valli, A., Dujovny, G. and García, J. A. (2008) Protease activity, self interaction, and small interfering RNA binding of the silencing suppressor p1b from *Cucumber vein yellowing ipomovirus*. *Journal of Virology,* **82,** 974-986.

Valli, A., Gallo, A., Calvo, M., Pérez, J. J. and García, J. A. (2014) A novel role of the potyviral helper component proteinase contributes to enhance the yield of viral particles. *Journal of Virology,* **88,** 9808-9818.

Valli, A., López-Moya, J. J. and García, J. A. (2007) Recombination and gene duplication in the evolutionary diversification of P1 proteins in the family *Potyviridae*. *Journal of General Virology,* **88,** 1016-1028.

Valli, A., Martín-Hernández, A. M., López-Moya, J. J. and García, J. A. (2006) RNA silencing suppression by a second copy of the P1 serine protease of *Cucumber vein yellowing ipomovirus*, a member of the family *Potyviridae* that lacks the cysteine protease HCPro. *Journal of Virology,* **80,** 10055-10063.

Valli, A. A., Gallo, A., Rodamilans, B., López-Moya, J. J. and García, J. A. (2018) The HCPro from the *Potyviridae* family: an enviable multitasking helper component that every virus would like to have. *Molecular Plant Pathology,* **19,** 744-763.

van Kasteren, P. B., Bailey-Elkin, B. A., James, T. W., Ninaber, D. K., Beugeling, C., Khajehpour, M., *et al.* (2013) Deubiquitinase function of arterivirus papain-like protease 2 suppresses the innate immune response in infected host cells. *Proceedings of the National Academy of Sciences USA,* **110,** 838-847.

Vasiljeva, L., Merits, A., Golubtsov, A., Sizemskaja, V., Kaariainen, L. and Ahola, T. (2003) Regulation of the sequential processing of *Semliki forest virus* replicase polyprotein. *Journal of Biological Chemistry,* **278,** 41636-41645.

Verchot, J. and Carrington, J. C. (1995a) Debilitation of plant potyvirus infectivity by P1 proteinase-inactivating mutations and restoration by second-site modifications. *Journal of Virology,* **69,** 1582-1590.

Verchot, J. and Carrington, J. C. (1995b) Evidence that the potyvirus P1 proteinase functions in trans as an accessory factor for genome amplification. *Journal of Virology,* **69,** 3668-3674.

Verchot, J., Herndon, K. L. and Carrington, J. C. (1992) Mutational analysis of the tobacco etch potyviral 35-kDa proteinase: identification of essential residues and requirements for autoproteolysis. *Virology,* **190,** 298-306.

Verdaguer, N., Ferrero, D. and Murthy, M. R. (2014) Viruses and viral proteins. *Journal from International Union of Crystallography ,* **1,** 492-504.

Vijayapalani, P., Maeshima, M., Nagasaki-Takekuchi, N. and Miller, W. A. (2012) Interaction of the *trans*-frame potyvirus protein P3N-PIPO with host protein PCaP1 facilitates potyvirus movement. *PLoS Pathogens,* **8,** e1002639.

Voinnet, O., Rivas, S., Mestre, P. and Baulcombe, D. (2003) An enhanced transient expression system in plants based on suppression of gene silencing by the p19 protein of *Tomato bushy stunt virus*. *Plant Journal,* **33,** 949-956.

Walsh, D., Mathews, M. B. and Mohr, I. (2013) Tinkering with translation: protein synthesis in virus-infected cells. *Cold Spring Harbor Perspectives in Biology,* **5,** a012351.

Wang, A., Carrier, K., Chisholm, J., Wieczorek, A., Huguenot, C. and Sanfaçon, H. (1999) Proteolytic processing of *Tomato ringspot nepovirus* 3C-like protease precursors: definition of the domains for the VPg, protease and putative RNA-dependent RNA polymerase. *Journal of General Virology,* **80** 799-809.

Wang, A. and Sanfaçon, H. (2000) Proteolytic processing at a novel cleavage site in the N-terminal region of the *Tomato ringspot nepovirus* RNA-1-encoded polyprotein *in vitro*. *Journal of General Virology,* **81,** 2771-2781.

Wang, D., Fang, L., Li, P., Sun, L., Fan, J., Zhang, Q.*, et al.* (2011a) The leader proteinase of *Foot-and-mouth disease virus* negatively regulates the type I interferon pathway by acting as a viral deubiquitinase. *Journal of Virology,* **85,** 3758-3766.

Wang, D., Fang, L., Liu, L., Zhong, H., Chen, Q., Luo, R.*, et al.* (2011b) *Foot-and-mouth disease virus* (FMDV) leader proteinase negatively regulates the porcine interferon-λ1 pathway. *Molecular Immunology,* **49,** 407-412.

Waterhouse, A. M., Procter, J. B., Martin, D. M., Clamp, M. and Barton, G. J. (2009) Jalview Version 2–a multiple sequence alignment editor and analysis workbench. *Bioinformatics,* **25,** 1189-1191.

Wei, T., Huang, T. S., McNeil, J., Laliberte, J. F., Hong, J., Nelson, R. S., *et al.* (2010) Sequential recruitment of the endoplasmic reticulum and chloroplasts for plant potyvirus replication. *Journal of Virology,* **84,** 799-809.

Weinheimer, I., Boonrod, K., Moser, M., Zwiebel, M., Fullgrabe, M., Krczal, G., *et al.* (2010) Analysis of an autoproteolytic activity of *Rice yellow mottle virus* silencing suppressor P1. *Biological Chemistry,* **391,** 271-281.

Wen, R., Zhang, S. C., Michaud, D. and Sanfacon, H. (2004) Inhibitory effects of cystatins on proteolytic activities of the *Plum pox potyvirus* cysteine proteinases. *Virus Research,* **105,** 175-182.

Whitfield, A. E., Ullman, D. E. and German, T. L. (2005) *Tomato spotted wilt virus* glycoprotein Gc is cleaved at acidic pH. *Virus Research,* **110,** 183-186.

Wilhelm, S. W., Bird, J. T., Bonifer, K. S., Calfee, B. C., Chen, T., Coy, S. R., *et al.* (2017) A student's guide to giant viruses infecting small eukaryotes: from *Acanthamoeba* to *Zooxanthellae*. *Viruses,* **17,** 46.

Willcocks, M. M., Brown, T. D., Madeley, C. R. and Carter, M. J. (1994) The complete sequence of a human astrovirus. *Journal of General Virology,* **75** 1785-1788.

Wright, D. A. and Voytas, D. F. (2002) *Athila 4* of *Arabidopsis* and *Calypso* of soybean define a lineage of endogenous plant retroviruses. *Genome Research,* **12,** 122-131.

Wu, C., Zhang, L., Li, P., Cai, Q., Peng, X., Yin, K., *et al.* (2016) Fragment-wise design of inhibitors to 3C proteinase from enterovirus 71. *Biochimica et Biophysica Acta,* **1860,** 1299-1307.

Wylie, S. J., Adams, M., Chalam, C., Kreuze, J., López-Moya, J. J., Ohshima, K., *et al.* (2017) ICTV virus taxonomy profile: *Potyviridae*. *Journal of General Virology,* **98,** 352-354.

Yost, S. A. and Marcotrigiano, J. (2013) Viral precursor polyproteins: keys of regulation from replication to maturation. *Current Opinion in Virology,* **3,** 137-142.

Young, B. A., Stenger, D. C., Qu, F., Morris, T. J., Tatineni, S. and French, R. (2012) *Tritimovirus* P1 functions as a suppressor of RNA silencing and an enhancer of disease symptoms. *Virus Research,* **163,** 672-677.

Zaitlin, M. (1998) The discovery of the causal agent of the tobacco mosaic disease. *Discoveries in Plant Biology*, Kung, S. D. and Yang, S. F. (eds) , pp 105-110.

APPENDIX

VII. Appendix

Exp. 1 band 1 - First Search

	Protein name	mRNA(cds)	Mass (Da)	Score	Matches
1	ubiquitin-specific protease 12	8285	132061	144	20
2	ubiquitin-specific protease 12	8286	131933	144	20
3	ubiquitin-specific protease 12	54097	131022	80	10
4	ubiquitin-specific protease 12	110332	130299	77	9
5	ubiquitin-specific protease 12	45188	131370	70	10
6	ubiquitin-specific protease 12	107798	131348	70	10
7	ubiquitin-specific protease 12	6538	110817	62	11
8	ubiquitin-specific protease 12	6540	110689	62	11
9	ubiquitin activating enzyme	46936	121070	55	13
10	ubiquitin activating enzyme	102877	121136	55	13
11	ubiquitin-specific protease 13	30244	29042	49	5
12	exoribonuclease 4	117627	39340	48	8
13	ribosomal protein L5	67163	34548	47	7
14	glutamate synthase 1	89147	171359	46	17
15	phosphoglycerate kinase	5087	42338	45	8
16	phosphoglycerate kinase	99217	42324	45	8
17	ubiquitin-specific protease 12	106279	19381	45	5
18	ascorbate peroxidase 6	3251	22460	44	6
19	ubiquitin-specific protease 12	90497	87151	44	10
20	Aldolase-type TIM barrel family protein	47073	45229	43	8

mRNA(cds): Accesion numbers as indicated in Sol Genomics Network (https://solgenomics.net/). Mass: Theoretical mass of the identified protein; Score: Based on Mascot; Matches: Total number of peptides identified.

Exp. 1 band 1 - Second Search

	Protein name	mRNA(cds)	Mass (Da)	Score	Matches
1	ubiquitin activating enzyme 2	46936	121070	58	11
2	ubiquitin activating enzyme 2	102877	121136	58	11
3	FUMARASE 2	29716	11849	48	5
4	chromomethylase 2	122996	45048	44	6
5	cysteine-rich RLK (RECEPTOR-like protein kinase) 8	48731	41635	43	6
6	alpha dioxygenase	132280	72924	43	9
7	ROP guanine nucleotide exchange factor 5	76041	35071	43	5
8	Minichromosome maintenance (MCM2/3/5) family protein41	8397	68905	41	9
9	ascorbate peroxidase 6	3251	22460	41	5
10	Minichromosome maintenance (MCM2/3/5) family protein41	113973	82553	40	9
11	Minichromosome maintenance (MCM2/3/5) family protein41	113974	82957	40	9
12	glutamate receptor 2.7	70421	68163	39	7
13	Minichromosome maintenance (MCM2/3/5) family protein41	113972	54486	39	8
14	Copper amine oxidase family protein	11868	85845	38	8
15	unknown protein	126848	24901	38	5
16	S-adenosyl-L-methionine-dependent methyltransferases superfamily protein	32788	69501	37	7
17	S-adenosyl-L-methionine-dependent methyltransferases superfamily protein	32789	69501	37	7
18	unknown protein	58275	17232	37	6
19	S-adenosyl-L-methionine-dependent methyltransferases superfamily protein	46163	69768	37	7
20	protein kinase 1B	104462	35038	37	6

mRNA(cds): Accesion numbers as indicated in Sol Genomics Network (https://solgenomics.net/). Mass: Theoretical mass of the identified protein; Score: Based on Mascot; Matches: Total number of peptides identified.

Exp. 1 band 2 - First Search

	Protein name	mRNA(cds)	Mass (Da)	Score	Matches
1	ubiquitin activating enzyme 2	46936	121070	484	40
2	ubiquitin activating enzyme 2	102877	121136	483	40
3	ubiquitin-specific protease 12	8285	132061	240	30
4	ubiquitin-specific protease 12	8286	131933	239	30
5	ubiquitin-specific protease 12	54097	131022	195	23
6	ubiquitin-specific protease 12	45188	131370	180	25
7	ubiquitin-specific protease 12	107798	131348	174	24
8	ubiquitin-specific protease 12	110332	130299	166	17
9	ubiquitin-specific protease 12	6538	110817	132	21
10	ubiquitin-specific protease 12	6540	110689	131	21
11	ubiquitin-specific protease 13	30244	29042	101	9
12	ubiquitin-specific protease 12	101410	51373	90	12
13	ubiquitin-specific protease 12	6539	61078	87	13
14	ubiquitin-specific protease 12	101409	43232	80	9
15	ubiquitin-specific protease 12	101408	39226	78	10
16	ubiquitin-specific protease 12	110333	44899	74	5
17	ubiquitin-specific protease 12	101407	31085	69	7
18	ubiquitin-specific protease 12	90497	87151	60	15
19	Pentatricopeptide repeat (PPR) superfamily protein	134552	89856	58	19
20	Pentatricopeptide repeat (PPR) superfamily protein	134553	89856	58	19

mRNA(cds): Accesion numbers as indicated in Sol Genomics Network (https://solgenomics.net/). Mass: Theoretical mass of the identified protein; Score: Based on Mascot; Matches: Total number of peptides identified.

Exp. 1 band 2 - Second Search

	Protein name	mRNA(cds)	Mass (Da)	Score	Matches
1	ubiquitin-specific protease 12	8285	132061	272	25
2	ubiquitin-specific protease 12	8286	131933	272	25
3	ubiquitin-specific protease 12	54097	131022	220	19
4	ubiquitin-specific protease 12	110332	130299	195	16
5	ubiquitin-specific protease 12	45188	131370	193	19
6	ubiquitin-specific protease 12	107798	131348	185	18
7	ubiquitin-specific protease 12	6538	110817	162	19
8	ubiquitin-specific protease 12	6540	110689	161	19
9	ubiquitin-specific protease 12	6539	61078	115	13
10	ubiquitin-specific protease 13	30244	29042	108	7
11	ubiquitin-specific protease 12	101410	51373	95	10
12	ubiquitin-specific protease 12	101409	43232	81	7
13	ubiquitin-specific protease 12	101408	39226	80	8
14	ubiquitin-specific protease 12	110333	44899	74	4
15	ubiquitin-specific protease 12	101407	31085	67	5
16	ubiquitin-specific protease 12	90497	87151	50	10
17	WWE protein-protein interaction domain protein family	112278	61839	50	10
18	Coatomer, alpha subunit	112063	77983	48	9
19	DNA repair-recombination protein (RAD50)	71212	109191	48	12
20	inflorescence deficient in abscission (IDA)-like 2	77506	13140	47	5

mRNA(cds): Accesion numbers as indicated in Sol Genomics Network (https://solgenomics.net/). Mass: Theoretical mass of the identified protein; Score: Based on Mascot; Matches: Total number of peptides identified.

Exp. 1 band 3 - First Search

	Protein name	mRNA(cds)	Mass (Da)	Score	Matches
1	ubiquitin activating enzyme 2	46936	121070	591	44
2	ubiquitin activating enzyme 2	102877	121136	590	44
3	ubiquitin-specific protease 12	8285	132061	71	21
4	ubiquitin-specific protease 12	8286	131933	71	21
5	Low-temperature-induced 65 kDa protein	115433	25096	60	8
6	ubiquitin-specific protease 12	101410	51373	54	10
7	annexin 5	105025	35800	53	10
8	glutathione peroxidase 4	79603	19255	46	8
9	glutathione peroxidase 4	79604	19255	46	8
10	ubiquitin activating enzyme 2	1997	122712	45	14
11	ubiquitin activating enzyme 2	1998	122712	45	14
12	Protein kinase superfamily protein	38771	61333	44	11
13	unknown protein	107605	104777	44	11
14	U5 small nuclear ribonucleoprotein helicase, putative	96046	77531	44	12
15	U5 small nuclear ribonucleoprotein helicase, putative	96047	77531	44	12
16	U5 small nuclear ribonucleoprotein helicase, putative	96048	77531	44	12
17	U5 small nuclear ribonucleoprotein helicase, putative	96049	77531	44	12
18	U5 small nuclear ribonucleoprotein helicase, putative	96050	77531	44	12
19	U5 small nuclear ribonucleoprotein helicase, putative	96051	77531	44	12
20	rabidopsis phospholipase-like protein (PEARLI 4) family	48389	27236	44	6

mRNA(cds): Accesion numbers as indicated in Sol Genomics Network (https://solgenomics.net/). Mass: Theoretical mass of the identified protein; Score: Based on Mascot; Matches: Total number of peptides identified.

Exp. 1 band 3 - Second Search

	Protein name	mRNA(cds)	Mass (Da)	Score	Matches
1	ubiquitin-specific protease 12	8285	132061	127	16
2	ubiquitin-specific protease 12	8286	131933	127	16
3	ubiquitin-specific protease 12	6538	110817	79	12
4	ubiquitin-specific protease 12	6540	110689	79	12
5	ubiquitin-specific protease 12	54097	131022	67	11
6	ubiquitin-specific protease 12	110332	130299	67	11
7	ubiquitin-specific protease 12	101410	51373	63	7
8	Low-temperature-induced 65 kDa protein	115433	25096	52	5
9	Phosphoribulokinase / Uridine kinase family	13389	50835	52	7
10	ubiquitin-specific protease 12	6539	61078	52	8
11	Phosphoribulokinase / Uridine kinase family	33956	31188	51	6
12	Phosphoribulokinase / Uridine kinase family	13388	32179	50	6
13	unknown protein	45489	13628	47	5
14	unknown protein	45492	13628	47	5
15	unknown protein	45495	13628	47	5
16	ubiquitin-specific protease 12	101409	43232	45	5
17	ascorbate peroxidase 3	3632	33498	44	5
18	S-adenosyl-L-methionine-dependent methyltransferases superfamily protein	1061	41227	42	6
19	ERD (early-responsive to dehydration stress) family protein	64099	18491	42	4
20	ubiquitin-specific protease 12	30246	53978	42	6

mRNA(cds): Accesion numbers as indicated in Sol Genomics Network (https://solgenomics.net/). Mass: Theoretical mass of the identified protein; Score: Based on Mascot; Matches: Total number of peptides identified.

Exp. 1 band 4 - First Search

	Protein name	mRNA(cds)	Mass (Da)	Score	Matches
1	ubiquitin activating enzyme 2	46936	121070	415	36
2	ubiquitin activating enzyme 2	102877	121136	415	36
3	unknown protein	57016	47034	50	11
4	gibberellin 2-oxidase	117138	37680	48	7
5	Leucine-rich repeat protein kinase family protein	71973	27481	46	7
6	Leucine-rich repeat protein kinase family protein	71975	27481	46	7
7	Pentatricopeptide repeat (PPR) superfamily protein	51314	16446	46	8
8	Pentatricopeptide repeat (PPR) superfamily protein	51316	16446	46	8
9	Pentatricopeptide repeat (PPR) superfamily protein	51317	16446	46	8
10	Spc97 / Spc98 family of spindle pole body (SBP) component	203	84175	46	11
11	basic helix-loop-helix (bHLH) DNA-binding superfamily protein	14734	40544	45	10
12	SET domain protein 16	51218	32352	45	7
13	SET domain protein 16	51219	32352	45	7
14	SET domain protein 16	51220	32352	45	7
15	SERINE-ARGININE PROTEIN 30	65828	28428	44	10
16	PRLI-interacting factor, putative	118133	59005	44	11
17	ARM repeat superfamily protein	38403	71101	43	13
18	Calcium-binding EF hand family protein	37096	79497	43	11
19	Calcium-binding EF hand family protein	37095	66080	43	10
20	YGGT family protein	97797	21086	43	7

mRNA(cds): Accesion numbers as indicated in Sol Genomics Network (https://solgenomics.net/). Mass: Theoretical mass of the identified protein; Score: Based on Mascot; Matches: Total number of peptides identified.

Exp. 1 band 5 - First Search

	Protein name	mRNA(cds)	Mass (Da)	Score	Matches
1	heat shock protein 70 (Hsp 70) family protein	103985	99484	311	31
2	heat shock protein 70 (Hsp 70) family protein	103984	88747	293	28
3	heat shock protein 70 (Hsp 70) family protein	119001	94924	261	31
4	Heat shock protein 70 (Hsp 70) family protein	120116	96951	178	29
5	Heat shock protein 70 (Hsp 70) family protein	120117	96951	178	29
6	Heat shock protein 70 (Hsp 70) family protein	120118	96951	178	29
7	Heat shock protein 70 (Hsp 70) family protein	120119	96951	178	29
8	Heat shock protein 70 (Hsp 70) family protein	120120	109083	164	29
9	Heat shock protein 70 (Hsp 70) family protein	56530	36360	152	16
10	Heat shock protein 70 (Hsp 70) family protein	56510	52316	132	13
11	Heat shock protein 70 (Hsp 70) family protein	56511	41735	118	10
12	Heat shock protein 70 (Hsp 70) family protein	134057	94516	100	18
13	heat shock protein 91	117463	94002	100	18
14	Heat shock protein 70 (Hsp 70) family protein	126045	17330	78	5
15	Heat shock protein 70 (Hsp 70) family protein	69282	17852	76	5
16	heat shock protein 91	68356	13430	66	9
17	Homeodomain-like transcriptional regulator	76185	133427	58	17
18	unknown protein	7005	126475	54	19
19	Mitochondrial substrate carrier family protein	53298	23894	52	9
20	arginine/serine-rich zinc knuckle-containing protein 33	76419	72639	51	15

mRNA(cds): Accesion numbers as indicated in Sol Genomics Network (https://solgenomics.net/). Mass: Theoretical mass of the identified protein; Score: Based on Mascot; Matches: Total number of peptides identified.

Exp. 1 band 5 - Second Search

	Protein name	mRNA(cds)	Mass (Da)	Score	Matches
1	Heat shock protein 70 (Hsp 70) family protein	119001	94924	254	24
2	Heat shock protein 70 (Hsp 70) family protein	120116	96951	170	22
3	Heat shock protein 70 (Hsp 70) family protein	120117	96951	170	22
4	Heat shock protein 70 (Hsp 70) family protein	120118	96951	170	22
5	Heat shock protein 70 (Hsp 70) family protein	120119	96951	170	22
6	Heat shock protein 70 (Hsp 70) family protein	120120	109083	155	22
7	heat shock protein 91	117463	94002	96	13
8	Heat shock protein 70 (Hsp 70) family protein	134057	94516	96	13
9	Heat shock protein 70 (Hsp 70) family protein	126045	17330	84	5
10	Heat shock protein 70 (Hsp 70) family protein	69282	17852	82	5
11	heat shock protein 91	68356	13430	67	8
12	heat shock protein 91	76185	133427	55	13
13	Mitochondrial substrate carrier family protein	53298	23894	55	8
14	unknown protein	125864	14099	52	7
15	unknown protein	7005	126475	52	15
16	Protein of unknown function (DUF1664)	86463	28055	50	7
17	sequence-specific DNA binding transcription factors	17363	61310	47	10
18	Exostosin family protein	8935	29626	46	8
19	unknown protein	120197	47164	44	9
20	Mitochondrial substrate carrier family protein	53297	33522	44	8

mRNA(cds): Accesion numbers as indicated in Sol Genomics Network (https://solgenomics.net/). Mass: Theoretical mass of the identified protein; Score: Based on Mascot; Matches: Total number of peptides identified.

Exp. 1 band 6 - First Search

	Protein name	mRNA(cds)	Mass (Da)	Score	Matches
1	TUDOR-SN protein 1	85936	107929	598	39
2	TUDOR-SN protein 1	56422	108005	547	37
3	TUDOR-SN protein 1	56424	108005	547	37
4	TUDOR-SN protein 1	56423	77916	370	32
5	heat shock protein 70 (Hsp 70) family protein	103985	99484	132	21
6	heat shock protein 70 (Hsp 70) family protein	103984	88747	123	18
7	tRNA synthetase class I (I, L, M and V) family protein	88669	67776	114	20
8	tRNA synthetase class I (I, L, M and V) family protein	56530	36360	114	12
9	heat shock protein 70 (Hsp 70) family protein	130747	11397	111	6
10	TUDOR-SN protein 1	52858	22342	107	8
11	tRNA synthetase class I (I, L, M and V) family protein	88668	67632	106	19
12	tRNA synthetase class I (I, L, M and V) family protein	124510	75327	81	18
13	TUDOR-SN protein 1	52854	71565	61	11
14	TUDOR-SN protein 1	52855	71565	61	11
15	CLPC homologue 1	108773	102453	61	22
16	GTP binding	131754	68150	60	16
17	Phosphatidylinositol 3- and 4-kinase family protein with FAT domain	60491	390806	58	36
18	tRNA synthetase class I (I, L, M and V) family protein	124511	108373	58	17
19	TUDOR-SN protein 1	130728	70360	57	10
20	Homeodomain-like transcriptional regulator	76185	133427	57	10

mRNA(cds): Accesion numbers as indicated in Sol Genomics Network (https://solgenomics.net/). Mass: Theoretical mass of the identified protein; Score: Based on Mascot; Matches: Total number of peptides identified.

Exp. 1 band 6 - Second Search

	Protein name	mRNA(cds)	Mass (Da)	Score	Matches
1	tRNA synthetase class I (I, L, M and V) family protein	88669	67776	149	19
2	tRNA synthetase class I (I, L, M and V) family protein	88668	67632	138	18
3	heat shock protein 70 (Hsp 70) family protein	103985	99484	129	15
4	heat shock protein 70 (Hsp 70) family protein	103984	88747	124	14
5	heat shock protein 70 (Hsp 70) family protein	56530	36360	112	9
6	tRNA synthetase class I (I, L, M and V) family protein	124510	75327	98	16
7	tRNA synthetase class I (I, L, M and V) family protein	124511	108373	74	15
8	RNA binding (RRM/RBD/RNP motifs) family protein	35983	96690	60	14
9	RNA binding (RRM/RBD/RNP motifs) family protein	35982	96777	60	14
10	Protein kinase superfamily protein	80331	47741	57	10
11	Protein kinase superfamily protein	123586	47869	57	10
12	heat shock protein 70 (Hsp 70) family protein	119001	94924	53	12
13	chloroplast beta-amylase	11103	61239	53	13
14	sec7 domain-containing protein	114401	130947	52	13
15	Protein kinase superfamily protein	13524	43404	52	8
16	histone mono-ubiquitination 2	58252	101226	51	14
17	ILITYHIA	50319	249911	51	18
18	senescence-related gene 1	17946	32266	48	7
19	chromatin remodeling factor18	71282	36215	46	8
20	Calcineurin-like metallo-phosphoesterase superfamily protein	53260	34974	46	7

mRNA(cds): Accesion numbers as indicated in Sol Genomics Network (https://solgenomics.net/). Mass: Theoretical mass of the identified protein; Score: Based on Mascot; Matches: Total number of peptides identified.

Exp. 1 band 6 - Third Search

	Protein name	mRNA(cds)	Mass (Da)	Score	Matches
1	heat shock protein 70 (Hsp 70) family protein	103985	99484	123	12
2	heat shock protein 70 (Hsp 70) family protein	103984	88747	115	11
3	heat shock protein 70 (Hsp 70) family protein	56530	36360	114	8
4	Protein kinase protein with tetratricopeptide repeat domain	47092	41198	55	7
5	Protein kinase protein with tetratricopeptide repeat domain	47093	41198	55	7
6	Transducin/WD40 repeat-like superfamily protein	119888	68658	50	8
7	Transducin/WD40 repeat-like superfamily protein	119889	68658	50	8
8	cell division cycle protein 48-related / CDC48-related	65590	102924	45	9
9	cell division cycle protein 48-related / CDC48-related	65593	102924	45	9
10	cell division cycle protein 48-related / CDC48-related	65591	132983	45	10
11	Plant protein of unknown function (DUF828)	64126	53609	45	8
12	Protein of unknown function (DUF3049)	104813	54716	44	7
13	Protein of unknown function (DUF3049)	104814	54716	44	7
14	sigma factor E	74920	59501	44	8
15	Transducin/WD40 repeat-like superfamily protein	119890	83767	43	8
16	ILITYHIA	50319	249911	43	13
17	quinolinate phoshoribosyltransferase	114110	20804	43	5
18	RECQ helicase L2	33078	17867	42	5
19	DegP protease 7	100967	84480	42	8
20	DegP protease 7	100962	62292	42	7

mRNA(cds): Accesion numbers as indicated in Sol Genomics Network (https://solgenomics.net/). Mass: Theoretical mass of the identified protein; Score: Based on Mascot; Matches: Total number of peptides identified.

Exp. 1 band 7 - First Search

	Protein name	mRNA(cds)	Mass (Da)	Score	Matches
1	TUDOR-SN protein 1	85036	107929	610	45
2	TUDOR-SN protein 1	56422	108005	560	43
3	TUDOR-SN protein 1	56424	108005	560	43
4	TUDOR-SN protein 1	56423	77916	418	37
5	tRNA synthetase class I (I, L, M and V) family protein	88669	67776	100	18
6	tRNA synthetase class I (I, L, M and V) family protein	88668	67632	92	17
7	TUDOR-SN protein 1	52854	71565	90	15
8	TUDOR-SN protein 1	52855	71565	90	15
9	TUDOR-SN protein 1	130728	70360	78	13
10	TUDOR-SN protein 2	52858	22342	73	7
11	tRNA synthetase class I (I, L, M and V) family protein	124510	75327	71	15
12	TUDOR-SN protein 1	130747	11397	65	4
13	tRNA synthetase class I (I, L, M and V) family protein	124511	108373	60	16
14	phragmoplast orienting kinesin 2	34945	240612	60	29
15	unknown protein	109062	119522	54	18
16	unknown protein	109063	119522	54	18
17	unknown protein	109065	119522	54	18
18	F-box family protein	97071	41343	52	9
19	phragmoplast orienting kinesin 2	34947	309999	51	30
20	unknown protein	78562	202966	50	22

mRNA(cds): Accesion numbers as indicated in Sol Genomics Network (https://solgenomics.net/). Mass: Theoretical mass of the identified protein; Score: Based on Mascot; Matches: Total number of peptides identified.

Exp. 1 band 7 - Second Search

	Protein name	mRNA(cds)	Mass (Da)	Score	Matches
1	tRNA synthetase class I (I, L, M and V) family protein	88669	67776	136	16
2	tRNA synthetase class I (I, L, M and V) family protein	88668	67632	123	15
3	tRNA synthetase class I (I, L, M and V) family protein	124510	75327	87	12
4	tRNA synthetase class I (I, L, M and V) family protein	124511	108373	60	11
5	exocyst complex component sec10	63936	43803	53	7
6	casein kinase 1	38344	53133	49	8
7	casein kinase 1	38345	53133	49	8
8	transmembrane receptors	37001	155996	48	12
9	Plant protein of unknown function (DUF936)	92938	56779	48	8
10	unknown protein	43096	33862	47	6
11	transmembrane receptors	13879	156403	47	12
12	exocyst complex component sec10	63935	58439	45	7
13	Protein kinase protein with adenine nucleotide alpha hydrolases-like domain	83931	77298	45	8
14	exocyst complex component sec10	63934	60389	44	7
15	Plant protein of unknown function (DUF247)	27265	12159	44	4
16	Agenet domain-containing protein	120981	42538	42	6
17	pre-mRNA-processing protein 40A	109763	63575	42	7
18	trithorax-like protein 2	104372	149516	41	11
19	Mitochondrial transcription termination factor family protein	55200	66880	40	7
20	Mitochondrial transcription termination factor family protein	55201	67966	40	7

mRNA(cds): Accesion numbers as indicated in Sol Genomics Network (https://solgenomics.net/). Mass: Theoretical mass of the identified protein; Score: Based on Mascot; Matches: Total number of peptides identified.

Exp. 1 band 8 - First Search

	Protein name	mRNA(cds)	Mass (Da)	Score	Matches
1	TUDOR-SN protein 1	85936	107929	610	38
2	TUDOR-SN protein 1	56422	108005	536	34
3	TUDOR-SN protein 1	56424	108005	536	34
4	TUDOR-SN protein 1	56423	77916	382	31
5	TUDOR-SN protein 2	52858	22342	108	7
6	TUDOR-SN protein 1	130747	11397	100	4
7	tRNA synthetase class I (I, L, M and V) family protein	88669	67776	83	15
8	tRNA synthetase class I (I, L, M and V) family protein	88668	67632	75	14
9	ubiquitin activating enzyme 2	46936	121070	67	18
10	ubiquitin activating enzyme 2	102877	121136	67	18
11	Homeodomain-like transcriptional regulator	76185	133427	54	16
12	NUP50 (Nucleoporin 50 kDa) protein	93803	47476	52	12
13	TUDOR-SN protein 1	52854	71565	50	10
14	TUDOR-SN protein 1	52855	71565	50	10
15	Cytochrome P450 superfamily protein	31340	40394	48	13
16	Homeodomain-like transcriptional regulator	76183	199051	47	19
17	casein kinase I	17653	28635	47	10
18	Radical SAM superfamily protein	97801	29112	47	8
19	phospholipase D beta 1	82082	96298	46	13
20	TUDOR-SN protein 1	130728	70360	46	9

mRNA(cds): Accesion numbers as indicated in Sol Genomics Network (https://solgenomics.net/). Mass: Theoretical mass of the identified protein; Score: Based on Mascot; Matches: Total number of peptides identified.

Exp. 1 band 8 - Second Search

	Protein name	mRNA(cds)	Mass (Da)	Score	Matches
1	tRNA synthetase class I (I, L, M and V) family protein	88669	67776	116	14
2	tRNA synthetase class I (I, L, M and V) family protein	88668	67632	104	13
3	ubiquitin activating enzyme 2	46936	121070	80	14
4	ubiquitin activating enzyme 2	102877	121136	80	14
5	exocyst complex component sec10	63936	43803	52	7
6	cell division cycle 48C	15897	90630	52	10
7	tRNA synthetase class I (I, L, M and V) family protein	124510	75327	52	9
8	casein kinase I	17653	28635	50	7
9	SET domain-containing protein	69157	17137	48	5
10	aminopeptidase M1	3671	64077	48	8
11	Homeodomain-like superfamily protein	59787	48882	48	7
12	unknown protein	92227	26564	47	6
13	unknown protein	92228	26564	47	6
14	FAR1-related sequence 3	11707	68911	47	8
15	FAR1-related sequence 3	11708	68911	47	8
16	FAR1-related sequence 3	11709	68911	47	8
17	FAR1-related sequence 3	11911	68911	47	8
18	FAR1-related sequence 3	17540	91545	47	9
19	tetratricopetide-repeat thioredoxin-like 3	126940	33603	46	6
20	uinone reductase family protein	53069	21608	45	5

mRNA(cds): Accesion numbers as indicated in Sol Genomics Network (https://solgenomics.net/). Mass: Theoretical mass of the identified protein; Score: Based on Mascot; Matches: Total number of peptides identified.

Exp. 2 Fraction 14

	Protein name	mRNA(cds)	Mass (Da)	Score	Matches	Sequences	emPAI
1	TUDOR-SN protein 1	85936	107929	2306	78(74)	25(24)	1.63
2	TUDOR-SN protein 1	56422	108005	1986	65(61)	22(21)	1.33
3	TUDOR-SN protein 1	52854	71565	853	32(29)	11(10)	0.83
4	chorismate synthase	3181	21059	656	24(24)	6(6)	2.37
5	chorismate synthase	3182	48275	946	33(33)	7(7)	0.87
6	ubiquitin-associated (UBA)/TS-N domain-containing protein	120910	47011	142	3(3)	2(2)	0.2
7	NADH-dependent glutamate synthase 1	22014	142729	494	20(17)	9(9)	0.32
8	NADH-dependent glutamate synthase 1	102188	116309	384	15(13)	15(13)	0.35
9	La protein 1	33365	51614	166	5(5)	3(3)	0.29
10	Heat shock protein 70 (Hsp 70) family protein	103984	8747	114	3(3)	1(1)	0.05
11	Heat shock protein 70 (Hsp 70) family protein	56510	52316	90	3(3)	1(1)	0.09
12	NADH-dependent glutamate synthase 1	29629	89686	85	3(3)	1(1)	0.05
13	ubiquitin-specific protease 12	45188	313370	278	11(9)	7(6)	0.22
14	ubiquitin-specific protease 12	8285/45188	132061	277	10(9)	6(6)	0.22
15	ubiquitin-specific protease 12	54097	131022	227	11(9)	8(7)	0.26
16	heat shock protein 89.1	20701	90201	500	18(17)	8(8)	0.47
17	lysyl-tRNA synthetase 1	38895	61879	27	1(1)	1(1)	*
18	unknown protein	28381	357094	52	4(2)	1(1)	0.13
19	Transducin/WD40 repeat-like superfamily protein	65900	134631	41	2(2)	1(1)	0.03
20	DEA(D/H)-box RNA helicase family protein	32753	58789	102	4(3)	1(1)	0.08
21	L-Aspartase-like family protein	46598	25013	95	3(3)	2(2)	0.41
22	eukaryotic translation initiation factor 2 gamma subunit	11051	51260	64	1(1)	1(1)	0.09
23	glycine decarboxylase P-protein 2	33746	114801	54	2(2)	2(2)	0.08
24	nitrogen fixation S (NIFS)-like 1	4210	50669	48	1(1)	1(1)	0.09
25	eukaryotic translation initiation factor 2 beta subunit	24094	30564	38	1(1)	1(1)	0.15
26	chromatin remodeling 42	22949	147766	36	3(2)	1(1)	0.03
27	Nodulin MtN3 family protein'	27879	29836	35	1(1)	1(1)	0.15
28	unknown protein	2683	253891	35	1(1)	1(1)	0.02
29	D-isomer specific 2-hydroxyacid dehydrogenase family protein	86325	34821	32	3(1)	1(1)	0.13
30	tryptophan synthase beta type 2	4649	54688	30	2(2)	1(1)	0.08
31	ucleoside diphosphate kinase family protein	48876	25739	28	2(0)	1(0)	*
32	forkhead-associated (FHA) domain-containing protein	134800	45489	27	1(1)	1(1)	0.1
33	ARM repeat superfamily protein	81166	126	26	1(0)	1(0)	*
34	global transcription factor group E8	58089	55080	27	1(1)	1(1)	0.08

mRNA(cds): Accesion numbers as indicated in Sol Genomics Network (https://solgenomics.net/). Mass: Theoretical mass of the identified protein; Score: Based on Mascot; Matches: Total number of peptides identified, in brackets, peptides with a score higher than the cutoff value; Sequences: Total number of unique sequences identified, in brackets, sequences with a score higher than the cutoff value, emPAI: exponentially modified protein abundance index, * no emPAI value was assigned.

Exp.3 Fraction GFC-cEF-14

	Protein name	mRNA(cds)	Mass (Da)	Score	Matches	Sequences	emPAI
1	TUDOR-SN protein 1	85936	107929	2944	104(93)	42(41)	4.35
2	TUDOR-SN protein 1	56422	108005	2373	87(75)	37(35)	3.2
3	TUDOR-SN protein 1	52854	71565	1060	43(37)	20(18)	1.95
4	TUDOR-SN protein 1	130728	70360	1017	41(36)	18(17)	1.82
5	chorismate synthase	3182	48275	896	35(31)	13(12)	1.9
6	Heat shock protein 70 (Hsp 70) family protein	119001	94924	789	25(23)	9(8)	0.5
7	heat shock protein 101	107329	101448	728	28(23)	13(12)	0.66
8	Heat shock protein 70 (Hsp 70) family protein	2408	73853	710	25(20)	12(10)	0.79
9	ubiquitin-associated (UBA)/TS-N domain-containing protein	120910	47011	676	19(17)	8(8)	1.07
10	Heat shock protein 70 (Hsp 70) family protein	94378	73664	663	23(20)	10(9)	0.69
11	NADH-dependent glutamate synthase 1	22014	142729	628	25(21)	19(16)	0.62
12	ubiquitin-associated (UBA)/TS-N domain-containing protein	127871	43864	628	17(15)	7(7)	0.98
13	NADH-dependent glutamate synthase 1	102188	116309	612	22(19)	17(15)	0.74
14	La protein 1	33365	51614	609	26(17)	14(11)	1.49
15	Heat shock protein 70 (Hsp 70) family protein	103985	99484	568	23(15)	11(6)	0.3
16	Heat shock protein 70 (Hsp 70) family protein	56510	52316	555	22(16)	8(5)	0.51
17	heat shock protein 91	117463	94002	554	17(15)	7(6)	0.32
18	heat shock protein 101	46625	101635	548	22(19)	13(12)	0.66
19	NADH-dependent glutamate synthase 1	29629	89686	473	17(14)	10(9)	0.54
20	glycine-rich protein	10661	48099	461	10(10)	5(5)	0.56
21	heat shock protein 70	15738	71459	436	12(11)	6(6)	0.43
22	heat shock cognate protein 70-1	12846	71321	415	12(11)	6(6)	0.44
23	ubiquitin-specific protease 12	8285	132061	385	18(12)	10(7)	0.26
24	ubiquitin-specific protease 12	6538	110817	365	17(12)	10(7)	0.31
25	Protein of unknown function (DUF3223)	111386	26062	362	8(8)	5(5)	1.26
26	glycine-rich protein	114497	43904	310	6(5)	2(2)	0.34
27	isovaleryl-CoA-dehydrogenase	105558	45476	260	9(6)	7(5)	0.6
28	ubiquitin-specific protease 12	54097	131022	223	10(7)	5(4)	0.14
29	rotamase FKBP 1	6740	63990	217	6(5)	4(3)	0.22
30	TUDOR-SN protein 2	52858	22343	196	9(7)	4(4)	1.14
31	glycine-rich protein	45863	47813	158	4(4)	2(2)	0.31
32	S-adenosyl-L-methionine-dependent methyltransferases superfamily protein	85755	42364	157	6(5)	5(4)	0.5
33	2-oxoglutarate dehydrogenase, E1 component	111073	116525	157	6(5)	5(5)	0.2
34	ARM repeat superfamily protein	114876	98817	154	5(4)	4(3)	0.14
35	ubiquitin-specific protease 12	90497	87151	152	7(6)	5(4)	0.22
36	CAP-binding protein 20	48309	29701	145	4(4)	1(1)	0.15
37	ribosomal RNA processing 4	15358	36733	142	6(5)	3(3)	0.42
38	Cleavage and polyadenylation specificity factor (CPSF) A subunit protein	47529	57766	135	4(3)	3(2)	0.16
39	Nucleic acid-binding, OB-fold-like protein	44788	21570	131	4(3)	3(2)	0.48
40	Cleavage and polyadenylation specificity factor (CPSF) A subunit protein	130628	66824	119	4(2)	3(2)	0.14
41	heat shock protein 89.1	20701	90201	110	5(4)	3(3)	0.15
42	SUMO-activating enzyme 2	16050	73337	108	3(3)	2(2)	0.12
43	Ribosomal protein S5 domain 2-like superfamily protein	117612	26942	102	3(3)	2(2)	0.37
44	pyruvate dehydrogenase E1 alpha	23594	48393	96	3(2)	2(1)	0.09
45	DEK domain-containing chromatin associated protein	64643	73772	91	3(2)	2(1)	0.06
46	S-adenosyl-L-methionine-dependent methyltransferases superfamily protein	85754	53779	91	3(2)	3(2)	0.17
47	Translin family protein	94536	17204	79	2(2)	2(2)	0.63

48	3'-5'-exoribonuclease family protein	91804	26879	79	3(2)	3(2)	0.37
49	Chaperone protein htpG family protein	51789	93163	76	3(2)	3(2)	0.1
50	PNAS-3 related	108808	22895	61	2(2)	1(1)	0.2
51	Adaptin family protein	94729	86250	61	1(1)	1(1)	0.05
52	phosphoenolpyruvate carboxykinase 1	77874	74082	58	1(1)	1(1)	0.06
53	3'-5'-exoribonuclease family protein	18720	24011	57	1(1)	1(1)	0.19
54	Small nuclear ribonucleoprotein family protein	76703	10815	54	1(1)	1(1)	0.47
55	tRNA synthetase class I (I, L, M and V) family protein	88668	67632	47	2(2)	1(1)	0.07
56	Protein kinase superfamily protein	25528	25215	46	1(1)	1(1)	0.18
57	methyl-CPG-binding domain 11	33535	35950	46	1(1)	1(1)	0.13
58	nucleolin like 2	58161	69890	44	1(1)	1(1)	0.06
59	ATP binding	104145	123643	43	2(0)	1(0)	*
60	Cytochrome P450 superfamily protein	7897	51680	43	1(1)	1(1)	0.09
61	3'-5'-exoribonuclease family protein	119444	30505	42	2(1)	2(1)	0.15
62	mitochondrial lipoamide dehydrogenase 1	21974	53692	41	1(1)	1(1)	0.08
63	unknown protein	28381	35794	40	2(2)	1(1)	0.13
64	cullin-associated and neddylation dissociated	982	86943	40	1(1)	1(1)	0.05
65	Mediator complex, subunit Med7	19486	19520	35	1(1)	1(1)	0.24
66	Transducin/WD40 repeat-like superfamily protein	92935	38247	34	1(1)	1(1)	0.12
67	Protein of unknown function, DUF647	107854	26543	33	1(1)	1(1)	0.17
68	actin 7	9884	41930	32	1(1)	1(1)	0.11
69	Eukaryotic translation initiation factor 2 subunit 1	82151	39205	31	1(0)	1(0)	*
70	minichromosome maintenance (MCM2/3/5) family protein	45834	94020	31	1(0)	1(0)	*
71	rhamnose biosynthesis 1	247	68209	31	1(0)	1(0)	*
72	Protein of unknown function (DUF1350)	121620	48252	29	1(1)	1(1)	0.09
73	unknown protein	109413	19005	27	1(0)	1(0)	*
74	methyltransferase 1	82133	178646	27	1(0)	1(0)	*
75	Putative adipose-regulatory protein (Seipin)	66203	62365	26	1(0)	1(0)	*
76	heat-shock protein 70T-2	59674	63134	25	1(0)	1(0)	*
77	heat shock protein 90.1	35038	81048	25	1(0)	1(0)	*

mRNA(cds): Accesion numbers as indicated in Sol Genomics Network (https://solgenomics.net/). Mass: Theoretical mass of the identified protein; Score: Based on Mascot; Matches: Total number of peptides identified, in brackets, peptides with a score higher than the cutoff value; Sequences: Total number of unique sequences identified, in brackets, sequences with a score higher than the cutoff value, emPAI: exponentially modified protein abundance index, * no emPAI value was assigned.

Exp. 3 Fraction GFC-cEF-11

	Protein name	mRNA(cds)	Mass (Da)	Score	Matches	Sequences	emPAI
1	Heat shock protein 70 (Hsp 70) family protein	119001	94924	864	30(27)	14(12)	0.79
2	Heat shock protein 70 (Hsp 70) family protein	94378	73664	586	17(16)	7(7)	0.5
3	NADH-dependent glutamate synthase 1	22013	107911	137	7(4)	7(4)	0.17
4	NADH-dependent glutamate synthase 1	102188	116309	155	6(5)	6(5)	0.2
5	La protein 1	33365	51614	100	3(3)	3(3)	0.28
6	Heat shock protein 70 (Hsp 70) family protein	103985	99484	837	30(24)	15(12)	0.67
7	Heat shock protein 70 (Hsp 70) family protein	56510	52316	7171	25(20)	11(10)	1.25
8	heat shock protein 91	117463	94002	517	17(15)	11(9)	0.5
9	NADH-dependent glutamate synthase 1	29629	89686	689	21(19)	9(8)	0.46
10	heat shock protein 70	15738	71459	424	11(10)	5(5)	0.35
a11	heat shock cognate protein 70-1	12925	71463	409	10(10)	5(5)	0.35
12	ribosomal RNA processing 4	15358	36733	145	9(9)	4(4)	0.58
13	Nucleic acid-binding, OB-fold-like protein	44788	21570	224	6(5)	3(3)	0.79
14	Nucleic acid-binding, OB-fold-like protein	50328	21777	187	5(5)	2(2)	0.47
15	Ribosomal protein S5 domain 2-like superfamily protein	117612	26942	118	5(4)	4(3)	0.6
16	Ribosomal protein S5 domain 2-like superfamily protein	3528	26954	63	3(2)	3(2)	0.37
17	Translin family protein	16151	31083	170	4(4)	2(2)	0.31
18	Translin family protein	94537	26750	418	10(8)	3(2)	0.87
19	3'-5'-exoribonuclease family protein	91804	26879	136	6(6)	4(4)	0.87
20	PNAS-3 related	108808	22895	87	4(4)	2(2)	0.44
21	Adaptin family protein	94729	86250	27	1(0)	1(0)	*
22	phosphoenolpyruvate carboxykinase 1	77874	74082	164	4(2)	3(1)	0.06
23	3'-5'-exoribonuclease family protein	18721	46313	74	3(2)	3(2)	0.2
24	3'-5'-exoribonuclease family protein	119444	30505	132	3(3)	2(2)	0.32
25	Transducin/WD40 repeat-like superfamily protein	92935	38247	52	2(2)	1(1)	0.12
26	Transducin/WD40 repeat-like superfamily protein	64331	36104	263	9(9)	6(6)	1.02
27	Transducin/WD40 repeat-like superfamily protein	46708	109011	39	2(1)	1(1)	0.04
28	Transducin/WD40 repeat-like superfamily protein	45834	94020	114	4(3)	3(3)	0.15
29	Transducin/WD40 repeat-like superfamily protein	31611	50557	50	1(1)	1(1)	0.09
30	Transducin/WD40 repeat-like superfamily protein	6439	108845	33	1(0)	1(0)	*
31	Transducin/WD40 repeat-like superfamily protein	29344	92660	31	2(0)	2(0)	*
32	DegP protease 7	51109	43365	339	9(8)	5(5)	0.63
33	translation initiation factor 3B1	66895	56407	332	11(10)	7(6)	0.57
34	DegP protease 7	18273	23771	319	7(6)	4(4)	1.03
35	DegP protease 7	131633	63143	232	8(6)	5(3)	0.22
36	eukaryotic translation initiation factor 3A	126735	108327	184	8(7)	4(3)	0.12
37	rhamnose biosynthesis 1	247	68209	139	6(6)	3(3)	0.21
38	eukaryotic translation initiation factor 3A	111605	111832	133	6(5)	4(3)	0.12
39	arginine methyltransferase 11	16984	35022	115	5(49	1(1)	0.27
40	Translation elongation factor EF1B/ribosomal protein S6 family protein	50919	25431	109	3(2)	2(2)	0.39
41	Ubiquitin-protein ligase Cullin 4	106876	86757	96	2(2)	2(2)	0.1
42	GTP binding Elongation factor Tu family protein'	18823	49601	89	2(2)	1(1)	0.09
43	cullin-associated and neddylation dissociated'	41849	136226	84	6(2)	5(2)	0.06
44	Cullin 1	28644	86865	80	2(2)	2(2)	0.1
45	phenylalanyl-tRNA synthetase, putative / phenylalanine--tRNA ligase	58452	56473	76	2(2)	2(2)	0.16
46	eukaryotic translation initiation factor 3G1	15484	32453	72	2(2)	1(1)	0.14
47	eukaryotic translation initiation factor 3E	13095	51431	70	2(2)	1(1)	0.09
48	splicing factor-related	15215	59970	59	2(2)	1(1)	0.07

49	NADP-malic enzyme 4	52564	70755	49	1(1)	1(1)	0.06
50	UBX domain-containing protein	30772	51696	47	2(2)	1(1)	0.09
51	2-cysteine peroxiredoxin B	46583	30047	42	1(1)	1(1)	0.15
52	Co-chaperone GrpE family protein	7594	33389	41	1(1)	1(1)	0.13
53	COP9 signalosome, subunit CSN8	18633	22684	37	1(1)	1(1)	0.1
54	eukaryotic translation initiation factor 3C'	29398	29398	37	1(1)	1(1)	0.05
55	cysteine-rich RLK (RECEPTOR-like protein kinase) 8	58805	55381	37	1(1)	1(1)	0.08
56	general regulatory factor 8	23462	28678	35	1(1)	1(1)	0.16
57	ATPase, V1 complex, subunit B protein	25797	61186	33	1(1)	1(1)	0.07
58	arginosuccinate synthase family'	36775	54364	33	1(1)	1(1)	0.08
59	translation initiation factor 3 subunit H1'	8216	39143	28	1(0)	1(0)	*
60	global transcription factor group E8	58089	55080	26	1(0)	1(0)	*

mRNA(cds): Accesion numbers as indicated in Sol Genomics Network (https://solgenomics.net/). Mass: Theoretical mass of the identified protein; Score: Based on Mascot; Matches: Total number of peptides identified, in brackets, peptides with a score higher than the cutoff value; Sequences: Total number of unique sequences identified, in brackets, sequences with a score higher than the cutoff value, emPAI: exponentially modified protein abundance index, * no emPAI value was assigned.